DOCUMENT RESUME

ED 434 140 TM 030 102

AUTHOR	Hamann, Thomas A.
TITLE	Evaluation of the 1996-97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle. Working Paper Series.
INSTITUTION	Bureau of the Census (DOC), Suitland, MD.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-WP-1999-03
PUB DATE	1999-02-00
NOTE	98p.
AVAILABLE FROM	U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Avenue, N.W., Room 400, Washington, DC 20208-5652; Tel: 202-219-1831.
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Evaluative (142)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Data Analysis; *Data Collection; Editing; Elementary Secondary Education; *Institutional Characteristics; *National Surveys; Program Evaluation; *Public Agencies; Schools; Tables (Data)
IDENTIFIERS	*Common Core of Data Program; National Center for Education Statistics

ABSTRACT

The Common Core of Data (CCD) program of the National Center for Education Statistics (NCES) consists of five separate surveys. An evaluation was conducted of the nonfiscal surveys from the CCD for 1996-97. It focused primarily on the agency and school universe surveys, but it also included the state nonfiscal survey. The evaluation explored such issues as respondent response time intervals, NCES/agent data processing and editing time intervals, and overall survey timeliness and data quality. The evaluation did not reveal any major problems or deficiencies in the 1996-97 nonfiscal CCD surveys data collection, processing, and editing cycle. Most of the recommendations resulting from the evaluation relate to improving overall survey timeliness, and only a few are directed toward the edits themselves. Most of the recommendations represent relatively minor changes to the procedures and processes of the existing system. Thirty-four principal findings and 18 principal recommendations are listed. Appendixes contain the agency universe and school universe error and warning message edit descriptions. (Contains 26 tables.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

NATIONAL CENTER FOR EDUCATION STATISTICS

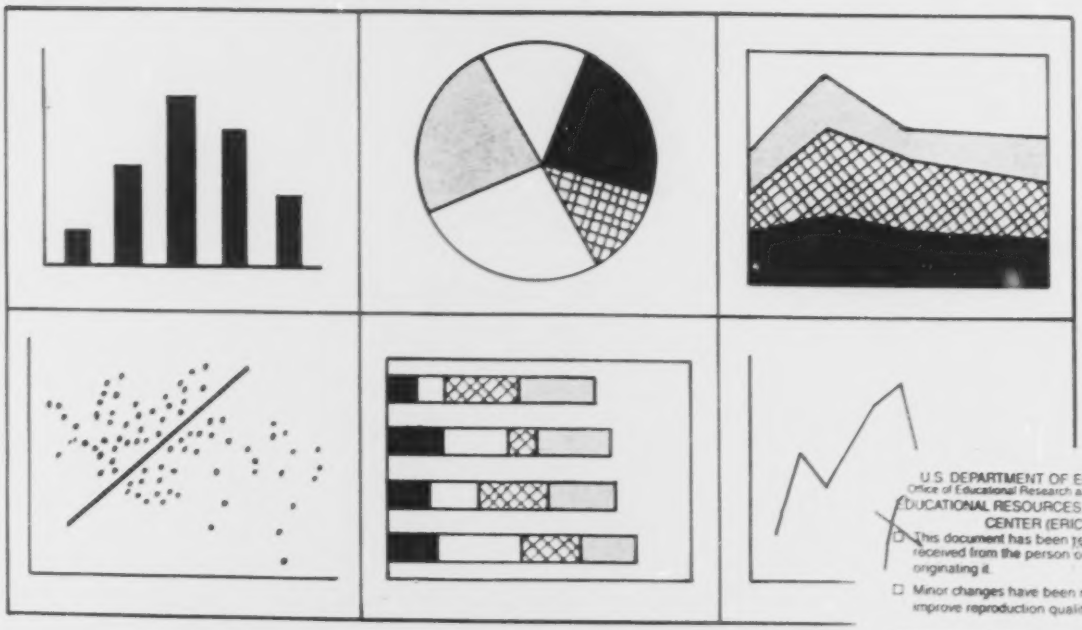
Working Paper Series

ED 434 140

Evaluation of the 1996-97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle

Working Paper No. 1999-03

February 1999



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM030102

U.S. Department of Education
Office of Educational Research and Improvement

BEST COPY AVAILABLE

Evaluation of the 1996-97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle

Working Paper No. 1999-03 February 1999

Contact: Beth Young
Elementary/Secondary and Library Studies Division
(202) 219-1562
e-mail: beth_young@ed.gov

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
C. Kent McGuire
Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208

The NCES World Wide Web Home Page is
http://nces.ed.gov

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *Evaluation of the 1996-97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle*. Working Paper No. 1999-03, by Thomas A. Hamann. Project Officer, Beth Young. Washington, D.C.: 1999.

February 1999

Foreword

In addition to official NCES publications, NCES staff and individuals commissioned by NCES produce preliminary research reports that include analyses of survey results, and presentations of technical, methodological, and statistical evaluation issues.

The ***Working Paper Series*** was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

To obtain copies of Working Papers please contact Angela Miles at (202)-219-1762, e-mail: angela_miles@ed.gov, or mail: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Ave. NW, Room 400, Washington, D.C. 20208-5654.

Marilyn M. McMillen
Chief Mathematical Statistician
Statistical Standards Program

Ralph Lee
Mathematical Statistician
Statistical Standards Program

Evaluation of the 1996-97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle

Prepared by:

Thomas A. Hamann
Bureau of the Census

Prepared for:

U.S. Department of Education
Office of Educational Research and Development
National Center for Education Statistics

February 1999

PURPOSE OF EVALUATION

This report was prepared for the National Center for Education Statistics (NCES) by the Governments Division of the U. S. Census Bureau. The principal author was Tom Hamann of the Governments Division. The report is part of a series that constitutes a comprehensive evaluation of the Common Core of Data (CCD) survey. The purpose of the series is to assess the quality of survey data as it relates to coverage, classification, processing, editing, reliability, and validity. It is intended to address issues of interest and concern to the NCES and the education community as a whole. This report contributes to the comprehensive evaluation by analyzing and providing the results of the evaluation of the data collection, processing, and editing cycle of the 1996-97 Nonfiscal CCD Surveys.

TABLE OF CONTENTS

FOREWORD iii

PURPOSE OF EVALUATION vi

LIST OF TABLES ix

ABBREVIATIONS USED IN THIS REPORT xi

CHAPTER 1. INTRODUCTION 1

Section 1.0 Survey Background and Purpose of Evaluation 1
Section 1.1 Principal Findings 1
Section 1.2 Principal Recommendations 5

CHAPTER 2. THE DATA COLLECTION, PROCESSING, AND EDITING CYCLE 9

Section 2.0 Introduction 9
Section 2.1 The 1996-97 CCD Data Collection Cycle 9
Section 2.2 The 1996-97 CCD Data Processing and Editing Cycle 10
Section 2.3 Timing Implications 11
Section 2.4 Impact of Resubmissions on Data Quality and Timeliness 12
Section 2.5 Recommendations 13

CHAPTER 3. AGENCY UNIVERSE AND SCHOOL UNIVERSE CCD SURVEYS PROCESSING AND EDITING STAGES 19

Section 3.0 Introduction 19
Section 3.1 Data Summary 19
Section 3.2 ID Assignment 20
Section 3.3 Internal Report 20
Section 3.4 Main Edit 21
Section 3.5 Post-Edit 22
Section 3.6 1996-97 CCD Data Correction and Response Rates 22
Section 3.7 Recommendations 24

CHAPTER 4. CCD MAIN EDIT PROGRAM EDITS AND RELATED ERROR AND WARNING MESSAGES 29

Section 4.0 Introduction 29
Section 4.1 Application of the Edits to Data Collection 30
Section 4.2 CCD Main Program Data Edits by Type 31
Section 4.3 Relational Edits 31
Section 4.4 Validation Edits 32
Section 4.5 Historical Edits 33
Section 4.6 Use of Edit System Software and 1996-97 CCD Data Error Rates 33
Section 4.7 Recommendations 34

CHAPTER 5. EFFECTIVENESS OF THE NONFISCAL CCD SURVEYS EDIT PROCESS 45

Section 5.0 Introduction 45
Section 5.1 Empirical Results of Initial Data and Final Data Comparison 45
Section 5.2 Implications for Individual CCD Nonfiscal Records 47
Section 5.3 Exploring Data Anomalies 48
Section 5.4 Tolerance/Range Implications for Selected Main Edit Program Edits 49
Section 5.5 Impact of Processing and Editing on Selected Final Data Category Counts 50
Section 5.6 Recommendations 51

CHAPTER 6. COMPARISON OF STATE NONFISCAL DATA TO AGENCY AND SCHOOL UNIVERSE DATA 61

Section 6.0 Introduction 61
Section 6.1 Cross-file Consistency 61
Section 6.2 Two-year Consistency 62
Section 6.3 Effectiveness of the CCD State Nonfiscal Survey Editing Process 62
Section 6.4 Recommendation 64

Appendix A. Education Agency Universe Error and Warning Messages/Edit Descriptions 69

Appendix B. School Universe Error and Warning Messages/Edit Descriptions 73

LIST OF TABLES

Table 1-1. Summary of Relevant Timings and Error/Warning Message Rates for 1996-97 Nonfiscal CCD Surveys 8

Table 2-1. Response Time Interval from Initial Mailout to Receipt of Agency, School, and State Nonfiscal Data Files in 1996-97 Nonfiscal CCD Surveys 14

Table 2-2. Summary of Initial Data File Submissions and Resubmissions by Month in 1996-97 Nonfiscal CCD Surveys 15

Table 2-3. Data Summary and Edit Phase Respondent Response Time Interval for Data Submission in 1996-97 Nonfiscal CCD Surveys 16

Table 2-4. Relevant Findings and Cycle Timings for 1996-97 Nonfiscal CCD Surveys 17

Table 2-5. Data Summary and Edit Phase Processing Time Interval for 1996-97 Nonfiscal CCD Surveys 18

Table 3-1. Summary of Number of Corrections Made at Data Summary and Edit Phases to Submitted Data Files in 1996-97 Nonfiscal CCD Surveys 26

Table 3-2. Summary of Total Number of Records in Error and Total Number of Errors Generated from Main Edit Program for Submitted Data Files in 1996-97 Nonfiscal CCD Surveys 27

Table 3-3. Summary of Percentage of Records in Error Reviewed and Errors Corrected by Respondents During Edit Phase in 1996-97 Nonfiscal CCD Surveys 28

Table 4-1. CCD Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD Agency Universe Survey 36

Table 4-2. CCD Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD School Universe Survey 39

Table 4-3. Main Edit Program Edits and Error and Warning Messages, by Source of Data Problem Detected, 1996-97 Nonfiscal CCD Surveys 41

Table 4-4. Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD Agency Survey, by Type of Edit 42

Table 4-5. Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD School Survey, by Type of Edit 44

Table 5-1. Error and Warning Messages Generated by Initial Submissions Versus Final Data Files, by State, 1996-97 CCD Agency Universe Survey 53

Table 5-2. Error and Warning Messages Generated by Initial Submissions Versus Final Data Files, by State, 1996-97 CCD School Universe Survey 54

Table 5-3. Error and Warning Messages Generated by Initial Submissions Versus Final Data Files, by Edit, 1996-97 CCD Agency Universe Survey 55

Table 5-4. Error and Warning Messages Generated by Initial Submissions Versus Final Data Files, by Edit, 1996-97 CCD School Universe Survey 56

Table 5-5. Most Frequently Generated Error and Warning Messages by Initial Data Submissions, 1996-97 Nonfiscal CCD Surveys 57

Table 5-6. Selected Error and Warning Messages Related to Blank and Missing Data Generated by Initial Data Submissions, 1996-97 Nonfiscal CCD Surveys 58

Table 5-7. Selected Error and Warning Messages with Range/Tolerance Criterion Generated by Initial Data Submissions, 1996-97 Nonfiscal CCD Surveys 59

Table 5-8. Comparison Between Initial Submissions and Final Data Files, by Category, for Selected Agency and School Total Data Counts, 1996-97 Nonfiscal CCD Surveys 60

Table 6-1. Comparison of State Nonfiscal to Agency and School Data Category Totals Generated by Initial Submissions Versus Final Date Files, by State, 1996-97 Nonfiscal CCD Surveys 65

Table 6-2. Comparison of State Nonfiscal to Agency and School Data Category Totals Generated by Initial Submissions Versus Final Data Files, by State, 1996-97 Nonfiscal CCD Surveys 66

Table 6-3. Comparison Between Initial Submissions and Final Data Files, by Category, for State Nonfiscal and Agency Data, 1996-97 Nonfiscal CCD Surveys 67

Table 6-4. Comparison Between Initial Submissions and Final Data Files, by Category, for State Nonfiscal and School Data, 1996-97 Nonfiscal CCD Surveys 68

ABBREVIATIONS USED IN THIS REPORT

The following abbreviations are used frequently throughout this report:

CCD—Common Core of Data

CMSA—Consolidated Metropolitan Statistical Area

CY—current year

DOD—Department of Defense

FIPS—Federal Information Processing Standards

FTE—full-time equivalency

ID—identification (number)

IEP—Individual Education Program

LEA—Local Education Agency

MSA—Metropolitan Statistical Area

na—not available

NCES—National Center for Education Statistics

NPEFS—National Public Education Fiscal Survey

PMSA—Primary Metropolitan Statistical Area

PY—prior year

SAS—computer statistical package:
computer language designed specifically for the manipulation of statistical data

(x)—not applicable

y—no response received

(z)—Entry would amount to less than half of the unit of measure shown

CHAPTER 1. INTRODUCTION

Section 1.0 Survey Background and Purpose of Evaluation

The Common Core of Data (CCD) program consists of five separate surveys. These include the "Public Elementary/Secondary Education Agency Universe Survey" (hereafter referred to as the agency survey), the "Public Elementary/Secondary School Universe Survey" (hereafter referred to as the school survey, the "State Nonfiscal Survey", the "National Public Education Finance Survey (NPEFS)", and the "Early Estimates Survey." The CCD contains three primary categories of information - identifying information, basic statistics, and fiscal data.

The National Center for Education Statistics' (NCES) goal for the CCD is a "comprehensive and timely national statistical database comparable across all states/territories on all public elementary and secondary schools, education agencies and programs." The purpose of the CCD is to provide basic statistical information on all children in this country receiving a free education from prekindergarten through grade twelve and on the public funds collected (revenues) and expended for providing free public elementary and secondary education.

The results of this evaluation will be used for ongoing process improvement of the CCD surveys. Findings from this survey may be used to improve the survey as a whole and to serve as an identification and basis for potential improvements in editing and processing of the nonfiscal components of the CCD surveys which are being expanded for the 1998-99 school year and collection cycle.

This evaluation explored such issues as respondent response time intervals, NCES/agent data processing and editing time intervals, as well as overall survey timeliness and data quality. A primary objective was to assess the effectiveness of the data editing procedures and activities that were part of the survey cycle. This evaluation consisted of a thorough review of the data edits, the associated error and warning messages, and the total and type of error counts generated by the submitted nonfiscal CCD data at various points in the 1996-97 survey cycle.

It is crucial to indicate at the outset that various terms used throughout this report, such as "errors," "edit failures," and "in-error records" do not necessarily always indicate incorrect data. The "errors" described herein represented situations which were flagged essentially because the data responses were not within expected ranges. The historical basis for the editing process has been primarily to alert respondents of potential data outliers or anomalies rather than to find data errors per se.

This evaluation covered only the nonfiscal surveys for the year 1996-97. It focused primarily on the agency and school universe surveys, but it also included the state nonfiscal survey. In this evaluation, the term "states" refers to all fifty states, the District of Columbia, the five outlying areas of American Samoa, Guam, the Commonwealth of Northern Mariana Islands, Puerto Rico, the U.S. Virgin Islands, and the U.S. Department of Defense (DOD) dependents overseas schools. Also, all totals and averages in the end-of-chapter tables include all the states and territories of the United States.

This evaluation did not reveal major problems or deficiencies in the 1996-97 Nonfiscal CCD Surveys data collection, processing, and editing cycle. Most of the recommendations provided herein relate to improving overall survey timeliness. Only a few are directed specifically towards the edits themselves. Many of the proposed suggestions, with few exceptions, represent relatively minor changes to the procedures and processes of the existing system. Indeed, some of the recommendations amount to mere presentation changes to the data reports which are generated and distributed to the states for review and response. Table 1-1 at the end of this chapter provides an overview of some of the more relevant findings of this evaluation.

Section 1.1 Principal Findings

The following highlight the main findings, by chapter, of this report. A more detailed discussion of these findings is presented in the subsequent chapters.

Chapter 2. The Data Collection, Processing, and Editing Cycle

1. The initial mailout for CCD survey year 1996-97 requesting data submission was about six weeks beyond the scheduled date for a majority of the states (some states' mailout were several months late). This delay occurred primarily as the direct result of unresolved issues with the previous year's data files.

2. The average state response time for initial data file submission of 15 weeks was about double the eight weeks allotted in the target schedule. The range of response time varied between two and 39 weeks.

Only two states submitted all three data files (agency, school, and state nonfiscal) by the due date, while 14 states took 30 or more weeks to do so.

3. There was a chronic "lateness" (time beyond due date), averaging about 14 weeks, associated with the states' initial data submissions. A full one-third of the respondents were as many as 16 weeks late in submitting their initial data files.

4. For two subsequent intervals, at the data summary and edit stages, in the survey cycle when states were asked to review and correct data, the respondents' average combined response time was 11 weeks (7 more weeks than the allocated 4 weeks). Thus, the amount of total time during the survey cycle attributable to late responses by the states averaged about 21 weeks.

5. The average total respondent response time was about 22 weeks, with a range of eight to 42 weeks.

6. The total survey cycle time, that is the time elapsed from initial mailout to final data file completion, for each state in the 1996-97 Nonfiscal CCD Surveys averaged just under one year at about 50 weeks. The longest total survey cycle time for any one state was 58 weeks while the shortest was 34 weeks.

7. During the reviewed survey cycle 17 states, nearly one-third of the respondents, provided either complete or partial revisions (some as many as three times) to their original data submissions.

8. As was the case for the previous year, these respondent response and survey completion timings had several consequences - the most compelling of which was to delay the subsequent year's initial mailout date and, hence, the entire survey cycle. Many of the issues relating to survey timeliness that were raised in this evaluation apparently would be addressed through a more timely initial mailout and completion of the annual survey cycle within the prescribed survey time frame.

Chapter 3. *Agency and School Universe CCD Surveys Processing and Editing Stages*

9. A major bottleneck and time delay in the processing and editing of nonfiscal CCD data files occurred as a result of issues and problems identified in the data summary (particularly the match report) and identification (ID) assignment stages of the survey cycle. Delays often occurred as a result of the NCES/agent having to wait for respondents to provide necessary data clarifications and corrections in order that they establish the overall survey universe and proceed with the data processing and editing.

10. In general, the internal report and the corresponding phase of the process and editing cycle were found to be thorough and useful, particularly in preparing the state data files for the more stringent edit checks to follow. However, many of the data checks and edits associated with the identifying and administrative information found in the internal report, while important for agency and school record completeness and accuracy, essentially performed validity checks on data which could have been conducted earlier in the process.

11. Review of the internal reports for all responding states revealed that 19 agency and 5 school data files generated no error messages at all. The edit messages associated with a variable name having three or fewer characters, such as school or agency name, city name, or street name, were reported some 304 times for the agency and school data. Yet manual review of these reports indicated that all of these three-letter variable names were valid, or at least, that they were not apparently changed or corrected.

12. The numerous reports generated by the post-edit appeared to identify issues which could, in many instances, have been resolved earlier in the processing and editing cycle. For the observed processing and editing cycle, apparently due to time and resource constraints, very little evaluative use was made of these reports.

13. The general overall "lateness" of the reviewed survey cycle, particularly as it affected the subsequent survey year's mailout, caused the processing for most states during the post-edit stage to be rather cursory. The primary focus at this phase in the survey cycle apparently was not one of edit, but simply to complete as final all remaining states' data files.

14. There was an alarmingly high non-response rate exhibited for both the data summary (30 percent) and the edit (65 percent) reports which were sent to the states for their review, correction, and

response. Only 20 states formally submitted initial data files, a data summary response, and an edit response.

15. For the 1996-97 Nonfiscal CCD Surveys, of the data files submitted by the state coordinators, just over 60 percent for the agency universe and about 30 percent for the school universe contained at least one error as generated by the CCD main edit program during the edit stage of the processing cycle.

16. The extent to which the data were reviewed and corrected throughout the survey cycle by the respondents was very minimal. Overall, less than one in five of all in-error records was sent to the respondents for review. Less than one-half of the respondents were provided the opportunity to review all of their agency records which generated errors during the edit phase, while only 20 percent of the respondents were provided this opportunity to review all of their error-filled school records. The number of identifiable corrections based on review of state responses for the edit stage revealed a minuscule overall survey average correction rate of 0.3 percent.

Chapter 4. *CCD Main Edit Program Edits and Related Error and Warning Messages*

17. The CCD main edit program edits performed three types of data checks: relational comparisons, data validation, and historical comparisons. The first group accounted for slightly more than half of all the generated error messages, while the validation edits represented more than 50 percent of the total edits themselves.

18. The agency and school universe data record files submitted by the state data coordinators for the 1996-97 Nonfiscal CCD Surveys contained a combined total of 72,819 error messages generated by 88 CCD main edit program edits. For the agency survey, 38 percent of the agency data records (agencies) generated no error messages while about 55 percent reported three or fewer. For the school survey, almost 70 percent of the schools generated no error messages while nearly all the remaining 30 percent generated 6 or fewer.

19. Close to 92 percent of the error/warning messages applied to the basic agency and school statistics themselves (i.e., variables associated with student and staffing counts), with two edits, those involving the comparison of current and prior year pupil/teacher ratio counts on both surveys, accounting for almost one-fifth of all error messages. Administrative information data elements (name, address, etc.) accounted for virtually all the remaining error messages, as those data elements involving the code classification-related edits (agency type or school operational status codes, for example) accounted for only between 0.3 and 0.4 percent of all errors.

20. In some cases, the error and warning messages tabulated and evaluated for this report were based on data submitted by the state CCD coordinators which had been previously processed through the Edit System software system provided to them to assist in their data reporting. For survey cycle 1996-97, thirty of 56 responding states used the provided Edit System software. Other states may have used their own software for such pre-submission editing purposes. Thus, in some instances, the evaluated data had already received at least one level of review prior to its submission. Consequently, it was not entirely possible to review the effectiveness of the CCD main edit program edits on the raw input data at the state level.

21. This evaluation did not find overwhelmingly strong evidence to support the assumption that the states which submitted data files after using the CCD Edit System software would have "cleaner" data files (i.e., more accurate and reliable data) and, hence, fewer errors and lower error rates. While the initial agency data files of states using the Edit System software exhibited about a two-thirds of an error per record less than those not using the system, the difference for the school data files was a negligible 0.01 of an error per record. The percentage of in-error records was slightly more than 32 percent for the states using the Edit System software versus 41 percent for those not using it.

Chapter 5. *Effectiveness of the Nonfiscal CCD Surveys Edit Process*

22. Comparing the number of CCD main edit program error/warning messages generated by the final data files to the initial files submitted by the states revealed only about a one-third overall reduction. The biggest decline in error and warning messages based on the final edits was in the historical edit category, which fell just over 40 percent. The smallest decline, for the agency data, was in relational edits, which dropped 15 percent. For the

school data, the validation edits, with a 25 percent reduction, exhibited the least decline.

23. For the agency survey, when evaluated by state, half the responses (28) showed a reduction in error counts between initial submitted and final data of less than 10 percent. Nine states reported no difference between the initial and final data. Two states accounted for more than half of the total reduction. For the school survey data, four states accounted for over three-fourths of the overall reduction in error messages. Six states reported an increase in error/warning messages.

24. When evaluated by edits, there were remarkable percentage differences between initial and final data files. For the surveys combined, the number of error messages generated by 35 (out of 82) edits declined by between 75 and 100 percent, 16 declined between 25 and 74.9 percent, and 31 declined less than 25 percent (including 2 that showed no change). Thirteen edits, however, actually generated more error messages from the final data files than from the initial data submissions.

25. The edits demonstrating the largest percent changes initially produced the fewest error messages. The number of error messages generated by edits whose percent change was greater than 50 percent accounted for only 15 percent of all error messages in the initial submissions. The 15 edits which had complete (100 percent) reduction in associated error messages accounted for less than 5 percent of the initial total. The 33 edits whose percent change was less than 25 percent accounted for nearly two-thirds of all messages in the initial data submissions.

26. Sixty-eight percent of the agency survey records and 37 percent of the school survey records initially submitted for the 1996-97 Nonfiscal CCD Surveys were in error. The vast majority (about 95 percent) of these records contained three or fewer error/warning messages. Only 17 records, out of a possible 104,831 records, contained more than 12 errors.

27. The 12 most error-generating edits for both the agency and school surveys accounted for nearly two-thirds of all initial data error messages generated. These edits reported a reduction in error messages between the initial and final data below the overall rate of 30 percent.

28. The 23 edits which generated error messages because data were not provided (potentially either blank, missing, or otherwise invalid) accounted for almost one-quarter of all errors found in the initial submissions. The edits which checked data to determine acceptability in terms of falling within a preselected range/tolerance accounted for 60 percent of the error messages. These edits demonstrated about a 25 percent reduction between initial and final data files.

29. Eight selected national-level data element final counts showed moderate change (none more than 15 percent) from initial data element counts.

Chapter 6. Comparison of State Nonfiscal Data to Agency and School Universe Data

30. The 13 data corrections made to the submitted data on the cross-file consistency report by the respondents represented a very small percentage (between 1 and 2 percent) of the data elements that were reviewed.

31. Respondent review and subsequent data editing for the CCD state nonfiscal data was quite limited. Of the 25 states for which verification of data was requested, only 48 verifiable data corrections or changes were made. Five states made no changes, while two states apparently did not respond to the request to verify their initial state nonfiscal data submission.

32. In general, the data figures reported for all three surveys - the agency, school, and state nonfiscal - demonstrated very little difference in total value. When evaluated by state, the comparison between state nonfiscal and agency universe data category average difference varied by about seven percent for the initial data submissions, down to near three percent for the final data files. Comparison between state nonfiscal data and school universe showed average state percent variation to be 2.62 percent for initial data, down a full one percent for the final data.

33. When evaluated by the data categories compared between the state nonfiscal and agency surveys, the final data totals showed a great deal of variation in the amount difference. Five of the 21 categories reported an improvement in data match of greater than 90 percent, while four data category totals compared between the two surveys showed increased variation between the initial data submissions and the final data files. For the state

nonfiscal and school surveys, all compared data categories demonstrated a reduction in the amount of difference between the two files. Five of 21 data categories had their total value come closer together by more than 70 percent.

34. Almost three-fourths of the variation or difference in data figures between the state nonfiscal survey and the two universe surveys was eliminated between the initial and final data files. This suggested reliable, comparable data across all three surveys.

Section 1.2 Principal Recommendations

Listed below is a summary of the primary recommendations specifically for the 1996-97 Nonfiscal CCD Surveys data collection, processing, and editing cycle. It is recognized that a few of the suggestions made herein, to varying degrees, are being considered for inclusion into the processing and editing rework currently taking place for the expanded CCD surveys in 1998-99.

1. Consideration should be given to the issue of the length of the CCD survey cycle time frame and the implications this has on the timeliness of the current year's completion and subsequent year's survey cycle commencement. All allotted response time interval lengths should be reviewed for sufficiency and appropriateness.

2. NCES or its agent should shorten the time taken to notify the state CCD coordinator that a data submission response has not been received by a certain scheduled due date. It seems crucial that non-response follow-up be improved. It is recommended that, since delinquent response contributes significantly to the overall delay in processing and given the very high non-response rates associated with these surveys, the state coordinators be notified immediately on a flow basis when a data response due date is past (or approaching). Such notification would likely have the effect of improving survey timeliness and data response rates as well.

3. Establish a concrete, adhered to "deadline" date for accepting initial data submissions and any resubmitted data. This would avoid an indefinite period of time in which data would be accepted. This event clearly extended the processing and overall survey cycle completion time of the reviewed cycle. This deadline date should be printed in the survey instructions manual which is part of the initial mailout materials and be part of any follow-up efforts to elicit state data response.

4. To improve (i.e., shorten) the length of the survey cycle, it is further recommended that collection agent's processing of submitted data be consistently initiated in a timely manner. This should be accomplished, to the extent possible, on a flow basis as the state submissions are received by NCES or its collection agent.

5. Given the very low survey follow up response and error-correction rates along with the condition that many error corrections were apparently initiated and addressed by NCES/agent, the efficacy of the existing post-submission edit and review process is brought into question. A radical approach to revising this process would involve revamping the edit cycle to eliminate (or significantly reduce), as much as possible, the procedure of returning edit materials/reports back to the states for their review. Such action would more formally place the burden and responsibility of data editing on NCES and its agent. This is to a large extent, however, what occurs anyway.

6. Other, less radical, recommendations would include adjustments to the current procedures. Several of these recommendations relate directly to the physical layout/presentation of the current data summary report. Alternative format suggestions include:

    a. Provide more clear and detailed descriptions of what is expected of the respondent in terms of reviewing and correcting the data presented in this report, including highlighting the due date, directly on the report itself.

    b. Delete the "five smallest and five largest values" headings and data for all data categories from the report.

    c. Place current year data and prior year data side-by-side and arrange dropout data totals in a tabular form (less the five smallest/largest data categories) to enhance ease of data comparison.

    d. For the school universe report, add a heading - "OPERATIONAL STATUS" - above the listing of schools (new, closed, etc.)

e. For the agency universe report, add a heading - "BOUNDARY STATUS" - above the listed LEA data categories.

7. More clearly highlight the allotted response time interval requested for state response, perhaps specifying an actual due date, in the instructions at the top of the match report.

8. Add an edit into the Edit System software program to compare assigned state and NCES identification (ID) numbers. This would require that the database listing of previously assigned ID numbers be provided with the initial reporting materials, but would eliminate the identification conflict "impossibilities" (such as a state education agency ID number being associated with more than one NCES education agency ID number on the school file) that appeared on the match report, often requiring resolution by the respondent and resulting in substantial processing time delays.

9. Provide respondents with capability, when necessary, to assign their own new ID numbers. This could, among other possibilities, be accomplished by:

   a. Modifying Edit System software to generate new NCES agency and school ID numbers from the ALLLEAID and ALLSCHNO data files (the database files from which new ID numbers are assigned) as states add new agencies or schools to their survey universes.

   b. Providing states some level of access to the ALLLEAID and ALLSCHNO data files directly via the internet.

10. Perform the internal report "edit" earlier in the editing and processing cycle. This would provide a barometer of the quality of a state's submission before further editing commences. Any potential data issues might be identified and addressed sooner in the process, perhaps included as part of the data summary program and report mailout stage, thereby eliminating the current practice of having to repeatedly contact respondents.

11. Modify the validity edits in the internal report program (similar to all subsequent edit programs) that examine identifying information data, like agency or school name, city name, or street name. It is recommended that the defining criteria be changed to identify responses having two characters or less rather than three characters or less. Although these certainly are not critical edits, this change would eliminate many of the error messages that state and federal reviewers and analysts must filter.

12. Replace the existing edit reports (which consist of the top 50 agency and top 100 school error- filled records) with an edit "summary report." It is recommended that such a report indicate the record (agency or school) ID number, the data element response in question, the error type (critical or warning), and the error message generated for all in-error records. This action would accomplish two important feats:

   a. Reduce the volume (and length) of edit materials requiring review that would be sent back out to the states, thus reducing respondent burden and likely improving response rates.

   b. Ensure that all in-error records (and individual errors) would be available for respondent review.

13. In order to make more productive use of the post-edit stage, consideration should be given to thoroughly reexamining the 28 generated post-edit summary reports for their intended purpose, function, and usefulness, as a component of the processing and editing cycle.

14. Effort should continue to be expended towards urging more state CCD coordinators to use the Edit System software and other internet functions made available to them for data collection and initial data file preparation and submission. Such effort could be made at the annual data conference training sessions, via the internet, and/or in communications (letter, telephone, fax, etc.) with individual state data coordinators. Based on the findings of this evaluation, at least moderate overall improvement in survey data quality could be realized if more states used the Edit System diskette when preparing and reporting their initial data submissions.

15. All edits (mostly relational) containing "acceptable" range/tolerance parameters (e.g., current year data value compared to prior year data value exceeded a ±25 percent difference) should be reevaluated for appropriateness and effectiveness. Some edit tolerances, for example, might include an absolute number change criterion or perhaps a threshold of comparison

could be employed (e.g., if a school has 25 or fewer students then some edits would be suppressed) for some edits. This would ensure a more realistic and useful data criterion "check" aimed at identifying genuine data errors and outliers while at the same time preventing excessive erroneous data failures from being generated. Specific recommendations for individual edits are found in Chapter 5.

16. As suggested in previous survey evaluations, consideration should be given to filling missing or erroneous ZIP code data using commercially available software. This function possibly could be included in the Edit System software revision planned for the upcoming CCD expansion. Although this data element is not crucial, this would eliminate the need for further reviewer and analyst attention.

17. Consideration should be given to adding a "remarks" field to the Edit System software and to state's data base file diskettes which are sent out in the initial mailout. Such a memo field would allow for the recording of pertinent comments and explanations regarding data features or anomalies. This would likely reduce the need for follow-up requiring contact by NCES or its collection agency with the respondent which inevitably lengthens the data processing time.

18. Incorporate the "State Nonfiscal Two Year Consistency" report and the "Cross-file Consistency" report into one report, highlighting the data which the respondents are requested to review or verify (all data exhibiting a 10 percent difference between prior and current years, for example). This action would not only reduce the redundancies of checking certain data more than once, but also would potentially reduce the number of times a state coordinator is queried about his/her data. A possible shortcoming is that such action still requires all three data files (state, school, and agency) to have been received and processed up to a certain point.

Table 1-1. Summary of Relevant Timings and Error/Warning Message Rates for 1996-97 Nonfiscal CCD Surveys

This table highlights several relevant processing and editing timings and error rates for data submitted in the 1996-97 Nonfiscal CCD Surveys. Note that all timings are in weeks.

State	Relevant Survey Timings*				Agency Universe#			School Universe#		
	Initial submission response	Lateness of initial submissions	Total cycle time	No. of times data resubmitted	% of in-error records	% of record errors corrected	% difference in no. of errors	% of in-error records	% of record errors corrected	% difference in no. of errors
Total/Average	15	14	50	28	62.4	0.3	-28.5	30.8	0.3	-30.2
Alabama	30	30	48	0	16.0	0.0	-9.8	13.3	0.0	-10.2
Alaska	21	18	50	3	43.6	-	-74.7	34.3	-	-5.8
Arizona	23	20	58	0	59.9	-	-25.9	24.7	-	-28.5
Arkansas	26	28	46	0	26.8	8.4	40.3	9.5	53.9	-58.9
California	12	14	50	1	75.0	-	-82.6	20.2	-	-8.3
Colorado	3	14	43	1	25.8	-	-14.5	21.0	-	-40.6
Connecticut	32	28	51	0	27.9	-	3.8	12.0	-	-16.5
Delaware	6	4	50	0	100.0	2.2	-50.0	23.9	3.9	-12.6
Dis. of Columbia	23	20	50	0	100.0	-	0.0	54.0	-	-2.0
Florida	6	3	50	0	36.5	-	-7.0	22.3	-	-16.7
Georgia	14	11	50	2	20.2	-	10.3	23.8	-	-8.7
Hawaii	35	32	50	0	100.0	-	-100.0	16.9	-	-26.3
Idaho	2	0	50	0	16.8	-	0.0	98.3	-	-3.7
Illinois	32	29	50	0	82.6	-	-2.3	14.1	-	-17.0
Indiana	15	15	48	0	19.6	-	-5.9	11.4	-	-5.7
Iowa	5	2	50	1	48.0	-	0.0	14.4	-	-1.8
Kansas	33	32	50	0	31.9	0.0	0.0	11.8	0.0	-7.4
Kentucky	34	31	52	0	100.0	-	-35.8	15.3	-	-41.2
Louisiana	25	22	52	3	44.4	-	-5.8	12.0	-	-8.2
Maine	4	2	50	0	67.2	0.0	-0.7	14.8	0.0	-2.4
Maryland	9	6	50	0	37.5	0.0	0.0	8.1	0.0	-18.8
Massachusetts	23	20	50	0	100.0	-	-23.5	98.7	-	-48.3
Michigan	16	33	34	3	100.0	-	-14.9	23.6	-	-24.9
Minnesota	30	33	52	1	86.1	-	-15.6	49.2	-	-15.8
Mississippi	9	6	50	0	48.8	2.8	-5.4	23.0	1.4	-9.8
Missouri	7	3	50	0	22.4	-	-13.1	18.3	-	-8.8
Montana	7	5	50	1	50.3	-	-5.0	25.9	-	-4.2
Nebraska	4	1	50	0	50.4	1.6	-11.7	36.4	0.0	-2.8
Nevada	8	4	50	0	38.9	0.0	-5.6	20.5	1.2	-5.1
New Hampshire	4	0	50	0	100.0	-	-22.7	15.9	-	-51.5
New Jersey**	-	-	-	-	-	-	-	-	-	-
New Mexico	6	3	50	0	27.0	32.1	-41.7	9.4	2.0	-2.0
New York	25	27	49	1	41.9	0.5	-78.1	13.0	-	-89.0
North Carolina	12	10	50	0	82.9	-	0.0	8.1	-	-11.8
North Dakota	4	1	50	0	25.6	0.0	-0.8	24.1	0.0	-20.9
Ohio	15	15	48	2	98.2	-	-25.8	15.8	-	-4.0
Oklahoma	26	29	46	0	99.8	-	-3.4	14.8	-	-9.4
Oregon	11	8	56	0	24.9	-	-44.9	14.4	-	-5.5
Pennsylvania	9	6	50	0	19.0	0.0	-1.4	9.8	0.0	-5.3
Rhode Island	14	11	52	2	21.6	0.0	0.0	22.5	1.0	-5.7
South Carolina	3	0	51	0	26.4	20.0	-25.0	15.6	1.1	-5.6
South Dakota	4	1	50	0	39.3	0.0	-1.7	31.8	0.0	-21.7
Tennessee	17	27	55	2	100.0	-	-30.4	98.9	-	-48.1
Texas	10	7	50	1	99.9	-	-7.9	98.1	-	-3.2
Utah	4	2	50	0	83.0	0.0	-22.8	10.3	3.0	-38.5
Vermont	36	33	50	0	39.9	-	-17.4	37.8	-	-21.2
Virginia	15	12	50	2	100.0	-	-25.7	98.1	-	-4.1
Washington	5	7	50	0	100.0	-	-10.6	25.1	-	2.1
West Virginia	13	10	50	0	31.6	0.0	-3.4	15.1	0.0	-39.6
Wisconsin	23	20	50	0	14.6	-	-16.2	11.2	-	-17.6
Wyoming	5	3	50	1	31.0	-	12.1	19.9	-	-12.6
Dept. of Defense	35	32	50	0	100.0	-	-38.6	95.4	-	-66.5
American Samoa	4	1	50	0	100.0	-	-100.0	100.0	-	-14.7
Guam	3	0	50	1	100.0	-	-93.5	45.7	-	42.1
North. Marianas	31	28	50	0	100.0	-	0.0	50.0	-	-8.1
Puerto Rico	12	15	45	0	100.0	100.0	-75.0	47.2	3.9	-80.7
Virgin Islands	4	1	50	0	100.0	0.0	0.0	14.3	27.3	-15.4

Notes: *The number of resubmissions is a total figure - all other figures in the first row are national averages.
#The percentage of in-error records and the percentage of record errors corrected are from submitted data at the edit stage of the cycle; the percent difference in number of generated error messages is the difference between the initial and final data files.
**New Jersey did not submit any data files for 1996-97.

BEST COPY AVAILABLE

CHAPTER 2. THE DATA COLLECTION, PROCESSING, AND EDITING CYCLE

Section 2.0 Introduction

The purpose of evaluating the data collection, processing, and editing cycle was to determine whether the chronology of the nonfiscal CCD surveys could have any impact on data quality and the overall survey timeliness. The scope of this evaluation was from initial "mailout" of survey materials through "closeout" and delivery of all states' final data files to NCES for CCD survey year 1996-97. No processing, editing, or respondent activity beyond 30 April 1998 was included in this report. This evaluation did not review any aspect of processing and editing done after delivery of final data files to the NCES. Nor did it review any associated final statistical products or publications of the survey.

The Nonfiscal Common Core of Data Surveys are conducted annually. The content of the surveys is determined by the National Center for Education Statistics (NCES) and the participants with a stated objective to report on the condition of public elementary and secondary education in the United States. Typically, the surveys are completed from administrative records of the state education agencies by state CCD coordinators who are designated by their chief state school officers as official state liaisons with the NCES.

The data collection proceeded via a mailout of the reporting instruction manual, software, and related data files to the state CCD coordinators, who in turn were responsible for the actual collection of data from their education agencies and schools.

Section 2.1 The 1996-97 CCD Data Collection Cycle

Review of recent survey years showed that the cycle for the 1996-97 nonfiscal surveys was not atypical of earlier censuses and that the data collection and processing was similar to the previous year's survey. The CCD reporting materials were sent to the state CCD coordinator in each state. The CCD survey materials included separate prior year survey data information for each of the three surveys provided in the states' preferred reporting format (electronic, shuttle, etc.), an edit software package designed to assist respondents with completing and editing of data files prior to initial submission, and a comprehensive reporting instructions manual. While the methods used by the states to collect their own education agency and school data vary, all survey responses must be certified by the state CCD coordinator as the responsible individual for confirming data availability and accuracy.

For the 1996-97 Nonfiscal CCD Surveys - agency, school, and state nonfiscal surveys - the targeted date for initial mailout was 15 January 1997. However, as the result of delays in completing the previous year's survey cycle, initial mailout did not commence until 23 February 1997 with mailing of materials to 43 states. The remaining 14 states were mailed initial materials periodically over the next several months as previous year's unresolved issues were resolved. The last initial mailout of materials did not occur until 5 June 1997. Table 2-1 shows actual mailout dates for all 57 states.

Each state submitted its agency, school, and state nonfiscal data to NCES or the collection agent in three separate, complete data files. State CCD coordinators had the option of submitting their data electronically on diskette or the Internet, or manually on a hard copy document referred to as a "shuttle". Shuttles were sent to state data coordinators for verification or correction of identifying or administrative information, as well as for the addition of new statistical data.

The stated due date for all three survey data files in the completion instruction manual for 1996-97 was 15 March 1997. This date was not modified in most cases as a result of the delayed initial mailout dates. Thus, for a majority of the states the actual time interval between the mailout of the reporting materials and the targeted due date was only about three weeks rather than the intended eight weeks. However, for those states which reporting materials were sent out to after the due date of 15 March, the revised due date was effectively five or six weeks after the actual date of mail out.

While the scheduled time interval between initial mailout and the data file due date of eight weeks represented a relatively short response time, it was not found to be a major concern for several reasons. First, the states collect basic elementary and secondary education statistics and information for their own purposes which meant that limited research was required to respond to the surveys. In some instances, however, the timing of this data collection may be long after CCD due dates[1]. Second, the reporting

[1] *NCES held a focus group discussion with several state CCD coordinators in July 1997 which focused on the issues of CCD timeliness and data submission deadlines. Some states indicated that they could not ensure*

requirements have up to this point remained rather consistent, changing only slightly, if at all, each year. Third, state CCD coordinators were typically notified well in advance of any anticipated change or modification in the data elements which were to be collected and reported.

When examining the nonfiscal CCD survey cycle for timing implications it was instructive to consider the respondent response time intervals for their initial response. That is, the time elapsed between initial mailout of reporting materials for the 1996-97 Nonfiscal CCD Surveys and submission of agency universe, school universe, and state nonfiscal data files by the states.

As the following chart and Table 2-2 at the end of this chapter indicate, there was a great deal of variance in the submission timings of initial data files, with only two states meeting the targeted due date:

CCD data files received by	Initial data submission		
	Agency	School	State non-fiscal
Total*	56	56	55
March 15, 1997**	2	2	2
March 31, 1997	11	10	10
April 30, 1997	12	12	10
May 31, 1997	5	5	4
June 30, 1997	7	7	4
July 31, 1997	4	4	5
August 31, 1997 . . .	2	3	5
September 30, 1997	4	4	5
October 31, 1997 . .	8	8	8
November 30, 1997	1	1	2
December 31, 1997	0	0	0

*New Jersey did not submit any data files for 1996-97; DOD did not submit a state nonfiscal data file.
**The "due" date for respondent data submissions.

LEAs reported data to them by the CCD due data of 15 March. Others indicated that their states have different due date timetables for various data elements.

Follow-up with the state coordinator regarding late data file submissions was somewhat limited. For late respondents, a letter or fax was sent, or telephone call was placed to inquire about the status of the state's submission as well as to remind them of the due date. However, as evident from review of the initial submission dates, there was a chronic lateness to most of these submissions.

For the reviewed survey cycle, the average response time for initial data file submission was about double the eight weeks allotted for in the target schedule. That is, about 15 weeks for the agency/school files (these were not considered submitted until both files had been received because existing procedures require both files for further processing and editing to occur) and about 16 weeks for the state nonfiscal data files. The range of response time for agency/school files varied between two and 36 weeks and between two and 39 weeks for the state nonfiscal data files. Less than half of the states, 21 and 20, respectively, for agency/school files and state nonfiscal files, submitted their initial data files within the allocated schedule time of eight weeks. Fourteen states took 30 or more weeks to initially submit all three data files. Table 2-1 shows initial response time intervals for all states.

Section 2.2 The 1996-97 CCD Data Processing and Editing Cycle

Data processing and editing refers to the combined activities of review, correction, and edit of the data files submitted by state CCD coordinators. These activities included the "processing" functions carried out by the NCES and/or its designated agent as well the "response" functions conducted by the respondent states.

For purposes of this evaluation, the agency universe and the school universe surveys, due to the inherent link between the data collected and reported for each survey, were for the most part evaluated together as a unit separate from the state nonfiscal survey. However, because of the very limited processing and editing done on the state nonfiscal data, this phase of the evaluation included that survey with the other two surveys. Also, since the focus of this chapter is on response and processing time intervals and their implications for data quality and overall survey timeliness, inclusion of the state nonfiscal seemed appropriate. Thus, the data processing and editing cycle for all three surveys will be reviewed in this section. More detailed coverage will appear in Chapters 3 and 4 for the agency and school

surveys, while Chapter 6 will expand on the state nonfiscal survey.

The processing and editing cycle for the 1996-97 CCD Agency and School Universe Surveys can be characterized as having five stages or phases. These stages, which to a large extent correspond to the SAS-based software programs used to review and edit data, include: data summary, ID assignment, internal report, main edit, and post-edit. While a detailed description and evaluation of these stages appear in the next chapter, the primary emphasis in this section was to assess the components of and lengths of identifiable time intervals which combined to make up the total processing and editing time cycle for the 1996-97 CCD Nonfiscal Surveys.

In addition to the initial mailout of reporting materials requesting initial state data submissions, there were two subsequent mailouts, at the data summary and edit phases in the cycle, which requested review and correction of previously submitted data on the part of the respondent. Table 2-3 shows the amount of time taken by the states to respond with comments or corrections at both the data summary and the edit stages. The average state response time to the data summary report was almost five weeks and for the edit report response, six weeks. In both instances, the states were requested to respond within two weeks. Note that the respondent response time interval for the data summary had a range of one day (typically a response by e-mail or fax) to 17 weeks. This interval for the edit stage ranged between one and 21 weeks.

These timings were critical to the overall process in that, in most cases, further processing activity was contingent upon receiving a response to these reports from the states. Stated differently, this time represented a period of time when the NCES or its agent were awaiting response and, thus, not fully capable of processing data files. Table 2-4 shows that the average total respondent response time, for the 20 states that submitted an initial data file, a data summary response, and an edit response, was about 22 weeks, with a range of 8 to 42 weeks.

The other timing component, internal NCES/agent processing time, of the overall processing and editing cycle is shown in Table 2-5. The intent of this table was to highlight two distinct components of the nonfiscal CCD survey cycle when, in practice, the burden of progressing towards completion of processing and editing activities lay with the NCES/agent. The first processing interval depicted the time between receipt of the agency and school data files (both of which must be received to generate the data summary report) and the mailout of the data summary report. For the cycle reviewed, the time taken to process the initial data files for data summary mailout took, on average, exactly nine weeks.

A second processing interval displayed in this table, which is not additive to the first, was the time taken from receipt of the state nonfiscal data file to mailout of the edit report. This mailing was not contingent upon receipt of a data summary report. However, the state nonfiscal data file must have been received in order to generate a cross-file consistency report which was mailed to the states as part of the edit report. The average time between these two points in the process was about 23 weeks.

It was clearly noted that in many instances these processing time intervals included time when the NCES or agent staff were awaiting a response to, or clarification of, an issue deemed critical for resumption of processing. This evaluation observed there to be no clear distinction between NCES/agent "processing and editing" time and the respondent "response" time. Thus, the total survey cycle time depicted in Table 2-4 cannot simply be divided into the two distinct time intervals of response and processing.

The important role and effect that the various respondent response times and internal processing time had on the overall length of the CCD survey cycle was, however, more evident. These timing implications are addressed in the following section.

Section 2.3 Timing Implications

This phase of the evaluation examined the data processing and editing cycle for the 1996-97 Nonfiscal CCD Surveys. This section focused on the surveys from the perspective of the total survey cycle completion time, including the processing/editing and response timings, and the resultant implications.

Table 2-4, among other relevant and important aspects of the reviewed surveys, shows that, on average, initial data file submission by the respondents was almost three and a half months, or 14 weeks, beyond the due date of 15 March 1997 regardless of the actual mailout date. In fact, slightly more than one-third of the respondents were as many as 4 months (16 weeks) late in submitting their initial agency or school universe data files. A timely survey cycle was apparently jeopardized from the beginning as a result of delinquent submission of initial data files by the states.

The total allocated respondent response time of four weeks (2 weeks for each phase) for the data summary and edit stages along with the combined average response time for these two phases of almost 11 weeks (shown in Table 2-3) meant an additional delay of 7 weeks. Thus, amount of time that was attributable to late responses by the states, averaged about 21 weeks[2] total when considering the 1996-97 Nonfiscal CCD Surveys scheduled time allocated for these activities. A question that needed to be asked: Were these scheduled response time intervals for these activities unreasonable, or could these five months of delays have been eliminated simply by states responding within the allotted time frame? Even if adjustments were made to lengthen allocated response-time intervals, there are apparent time savings associated with response timings such as electronic follow-up responses by the states which could shave several weeks from the overall response time. In any event, the issue was critical to a timely survey cycle.

Table 2-4 reveals that the total survey cycle time, that is the time elapsed from initial mailout to final data file completion, for each state in the 1996-97 Nonfiscal CCD Surveys averaged just under 1 year at about 50 weeks. However, the elapsed time from the first states' initial reporting material mailout to the last states' data file being finalized - the longest total survey cycle time for any one state - was 58 weeks. The shortest interval for any state was 34 weeks.

In terms of evaluating the overall survey processing and editing cycle duration (total time from initial mailout to complete final data file closeout) for the 1996-97 Nonfiscal CCD Surveys, the response and processing timings discussed in this chapter had several consequences:

1. Delayed the commencement of initial data file processing and editing.

2. Compressed the time available for data review and correction, particularly in the latter stages of the cycle such as the post-edit stage.

3. State data coordinators were queried about their data, in most cases, many weeks after the original submissions. This practice had the potential for making follow-up inquiries more difficult because of excessive time lapse.

4. Delayed public release file of data and NCES publications that use the data.

5. Ultimately, delayed the subsequent year's mailout date and entire survey cycle.

Section 2.4 Impact of Resubmissions on Data Quality and Timeliness

Another remarkable aspect of this cycle was the large number of states that resubmitted data at some point during the processing and editing cycle. During the evaluated survey cycle, 17 states provided 27, either partial or complete, revisions to their data. For the purposes of this evaluation, if a state resubmitted an entire agency or school data file, the "resubmission" was considered complete. Those resubmitted data which consisted of data related to a specific category, like dropout or staffing data for example, were considered to be partial resubmissions

As Table 2-2 indicates, about half of the states that resubmitted did so more than once, some as many as three times during the survey cycle. The resubmissions were generally spread throughout the entire survey cycle and covered about one full year, with one state resubmitting data as early as April 1997 and one state as late as April 1998. While this table does not necessarily address the additional processing that was required due to resubmitted data files, it does demonstrate the rather significant amount of data that required some level of re-processing and re-editing. The number and frequency of data resubmissions carried both data quality and survey timeliness implications.

Twelve of the 17 states that resubmitted data took longer than the national average of 49.8 weeks in total survey cycle completion time. The average total processing completion time for those states that resubmitted versus those that did not was 49.1 weeks compared to 50.2 weeks. While this one week differential is certainly not conclusive evidence of a detrimental impact of data resubmission during the nonfiscal CCD survey cycle, the issue of data resubmission, nonetheless, seemed to be relevant when considering survey cycle completion time.

The necessity of having to re-process the same (revised) data more than once clearly contributed to redundancies and an overall diversion of effort and resources that otherwise might have been applied to further processing of other states' data. Simply stated, time and resources spent on data resubmissions were essentially time and

[2] *Included the almost 14 weeks attributable to late initial data file submissions and the 7 weeks, 3 weeks for the data summary response and 4 weeks for the edit response, associated with late state responses.*

resources not spent on processing other data and undoubtedly lengthened survey processing time.

Section 2.5 Recommendations

There are several recommendations associated with the response and processing timings of the data collection, processing, and editing described herein:

1. Careful consideration should be given to the issue of the length of the CCD survey cycle time frame and the implications this has on the timeliness of the current year's completion and subsequent year's survey cycle commencement.

2. NCES or its agent should shorten the time taken to notify the state CCD coordinator that a data submission response has or has not been received by a certain scheduled due date. It is recommended that, since delinquent respondent responses contribute significantly to the overall delay in processing, the states be notified immediately on a flow basis when a data due date is past (or approaching). Such notification would likely have the effect of improving survey data request response rates as well.

3. Establish a concrete, adhered to "deadline" date for accepting initial data submissions and any resubmitted data. This would avoid an indefinite period of time in which data would be accepted. This event clearly extended the processing and overall survey cycle completion time. This deadline date should be printed in the survey instructions manual which is part of the initial mailout materials.

4. To improve (i.e., shorten) the length of the survey cycle, it is further recommended that collection agent's processing of submitted data be consistently initiated in a timely manner. This should be accomplished, to the extent possible, on a flow basis as the state submissions are received by NCES or its collection agent.

Table 2-1. Response Time Interval from Initial Mailout to Receipt of Agency, School, and State Nonfiscal Data Files in 1996-97 Nonfiscal CCD Surveys

This table shows the interval (in weeks) between initial mailout and the date the various data files were received. That is, it shows the initial response time interval. Note that the due date for data submission was 15 March 1997.

State	Initial Mailout	Agency Received	School Received	State Non-fiscal Rec'd.	Response Interval (in weeks) Agency/ School*	State Nonfiscal
All states, avg.	-	-	-	-	15	16
Alabama	3/12/97	4/01/97	10/08/97	10/08/97	30	30
Alaska	2/23/97	7/17/97	7/17/97	7/17/97	21	21
Arizona	2/23/97	7/30/97	7/30/97	7/30/97	23	23
Arkansas	4/02/97	9/30/97	9/30/97	9/30/97	26	26
California	4/02/97	6/20/97	6/10/97	7/23/97	12	17
Colorado	6/05/97	6/23/97	6/23/97	6/24/97	3	3
Connecticut	2/23/97	9/30/97	4/28/97	10/24/97	32	35
Delaware	2/23/97	4/09/97	4/09/97	4/07/97	6	6
Dis. of Columbia	2/23/97	7/30/97	7/30/97	7/30/97	23	23
Florida	2/23/97	4/07/97	4/07/97	4/07/97	6	6
Georgia	2/23/97	5/30/97	5/30/97	5/30/97	14	14
Hawaii	2/23/97	10/23/97	9/18/97	9/18/97	35	30
Idaho	2/23/97	3/10/97	3/10/97	8/04/97	2	23
Illinois	2/23/97	10/02/97	10/02/97	11/19/97	32	39
Indiana	3/12/97	6/25/97	6/25/97	10/30/97	15	33
Iowa	2/23/97	3/31/97	3/31/97	3/31/97	5	5
Kansas	3/12/97	10/24/97	10/24/97	10/23/97	33	33
Kentucky	2/23/97	10/20/97	10/20/97	8/27/97	34	27
Louisiana	2/23/97	8/14/97	8/14/97	10/30/97	25	35
Maine	2/23/97	3/25/97	3/25/97	3/25/97	4	4
Maryland	2/23/97	4/23/97	4/23/97	4/23/97	9	9
Massachusetts	2/23/97	7/31/97	7/31/97	10/29/97	23	36
Michigan	7/14/97	11/01/97	11/01/97	11/13/97	16	18
Minnesota	4/02/97	10/31/97	10/31/97	7/30/97	30	17
Mississippi	2/23/97	4/28/97	4/28/97	4/28/97	9	9
Missouri	2/23/97	4/08/97	4/08/97	4/16/97	7	8
Montana	2/23/97	4/18/97	4/18/97	8/04/97	7	23
Nebraska	2/23/97	3/24/97	3/24/97	3/24/97	4	4
Nevada	2/23/97	4/15/97	4/15/97	4/15/97	8	8
New Hampshire	2/23/97	3/18/97	3/18/97	3/18/97	4	4
New Jersey**	4/03/97	-	-	-	-	-
New Mexico	2/23/97	4/07/97	4/07/97	4/23/97	6	8
New York	4/03/97	9/23/97	9/23/97	9/23/97	25	25
North Carolina	2/23/97	5/19/97	5/19/97	5/20/97	12	12
North Dakota	2/23/97	3/25/97	3/25/97	3/25/97	4	4
Ohio	3/12/97	6/24/97	6/24/97	8/12/97	15	22
Oklahoma	4/03/97	10/01/97	10/01/97	10/28/97	26	29
Oregon	2/23/97	5/12/97	5/12/97	3/06/97	11	2
Pennsylvania	2/23/97	4/23/97	4/23/97	4/23/97	9	9
Rhode Island	2/23/97	6/03/97	6/03/97	6/03/97	14	14
South Carolina	2/23/97	3/14/97	3/14/97	3/14/97	3	3
South Dakota	2/23/97	3/19/97	3/19/97	3/19/97	4	4
Tennessee	2/23/97	9/19/97	8/14/97	9/24/97	17	18
Texas	2/23/97	4/30/97	4/30/97	4/30/97	10	10
Utah	2/23/97	3/26/97	3/26/97	4/09/97	4	7
Vermont	2/23/97	10/31/97	10/31/97	9/04/97	36	28
Virginia	2/23/97	6/05/97	6/05/97	10/27/97	15	35
Washington	4/03/97	5/05/97	5/05/97	5/19/97	5	7
West Virginia	2/23/97	5/21/97	5/21/97	5/21/97	13	13
Wisconsin	2/23/97	8/04/97	8/04/97	8/04/97	23	23
Wyoming	2/23/97	4/01/97	4/01/97	6/10/97	5	15
Dept. of Defense**	2/23/97	10/24/97	10/24/97	-	35	-
American Samoa	2/23/97	3/21/97	3/21/97	3/21/97	4	4
Guam	2/23/97	3/18/97	3/18/97	3/18/97	3	3
North. Marianas	2/23/97	3/18/97	9/24/97	3/18/97	31	3
Puerto Rico	4/03/97	6/23/97	6/23/97	6/23/97	12	12
Virgin Islands	2/23/97	3/24/97	3/24/97	3/24/97	4	4

Notes: *Response time interval is from initial mailout to the time of receipt of the latest file (agency or school).
**New Jersey did not submit any data files for 1996-97; DoD did not submit a state nonfiscal data file.

Table 2-2. Summary of Initial Data File Submissions and Resubmissions by Month in 1996-97 Nonfiscal CCD Surveys

This table shows the distribution by month of initial data file submissions and resubmissions during the 1996-97 Nonfiscal CCD Surveys data processing and editing cycle.

State	Month of Initial Data Submission 1997									Month of Data Resubmission 1997/1998											
	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Total*	37	34	14	18	13	10	13	24	4	1	1	1	4	1	4	3	4	3	3	2	0	1
Alabama						A		†														
Alaska					✓										C			C				P
Arizona					✓																	
Arkansas							✓															
California				×	St								C									
Colorado				✓									C									
Connecticut		S					A	St														
Delaware		✓																				
Dis. of Columbia					✓																	
Florida		✓																				
Georgia			✓														C	C				
Hawaii							†	A														
Idaho	×					St																
Illinois								×	St													
Indiana				×				St														
Iowa	✓									C												
Kansas								✓									—					
Kentucky						St		×														
Louisiana						×		St								C			C	C		
Maine	✓																					
Maryland		✓																				
Massachusetts					×			St														
Michigan									✓								C		C	C		
Minnesota					St			×									C					
Mississippi		✓																				
Missouri		✓																				
Montana		×				St									P							
Nebraska	✓																					
Nevada		✓																				
New Hampshire	✓																					
New Jersey**																						
New Mexico		✓																				
New York							✓								P							
North Carolina			✓																			
North Dakota	✓																					
Ohio				×		St							C	C								
Oklahoma								✓														
Oregon	St		×																			
Pennsylvania		✓																				
Rhode Island				✓														C	C			
South Carolina	✓																					
South Dakota	✓																					
Tennessee						S	□									C	P					
Texas		✓														P						
Utah	×	St																				
Vermont							St	×														
Virginia				×				St					P		P							
Washington	✓																					
West Virginia			✓																			
Wisconsin						✓																
Wyoming		×		St							C											
Dept. of Defense								×														
American Samoa	✓																					
Guam	✓											C										
North. Marianas	□						S															
Puerto Rico				✓																		
Virgin Islands	✓																					

Notes: *Indicates the total number of data files (agency, school, or state nonfiscal), or partial/complete resubmissions, that were submitted.
**New Jersey did not submit any data files for 1996-97.

Meaning of symbols: A = Agency data file only; S = School data file only; St = State nonfiscal data file only; ✓ = All three data files; × = Agency and School data files; † = School and State nonfiscal data files; □ = Agency and State data files; C = complete resubmission; P = partial resubmission

BEST COPY AVAILABLE

Table 2-3. Data Summary and Edit Phase Respondent Response Time Interval for Data Submission in 1996-97 Nonfiscal CCD Surveys

This table shows the interval between mailout and receipt of data summary and edit reports. That is, it shows the respondent response time interval for the data summary and edit phases of the CCD processing cycle.

State	Data Summary Mailed	Data Sum. Response Received	Response Interval (in weeks)*	Edit Mailed	Edit Response Received	Response Interval (in weeks)*
All states, avg.	-	-	5	-	-	6
Alabama	10/17/97	10/31/97	2	10/24/97	11/03/97	1
Alaska	11/18/97	1/05/98	7	12/17/97	y	-
Arizona	8/06/97	y	-	4/02/98	y	-
Arkansas	10/22/97	12/04/97	6	1/27/98	2/17/98	3
California	7/08/97	7/23/97	2	3/24/98	y	-
Colorado	2/24/98	2/25/98	1 day	4/02/98	y	-
Connecticut	11/04/97	11/14/97	2	12/17/97	y	-
Delaware	7/08/97	9/25/97	11	11/25/97	12/05/97	2
Dis. of Columbia	11/20/97	y	-	12/17/97	y	-
Florida	5/20/97	6/20/97	4	1/08/98	y	-
Georgia	7/03/97	9/09/97	9	9/22/97	y	-
Hawaii	11/19/97	y	-	12/17/97	y	-
Idaho	5/20/97	6/20/97	4	9/22/97	y	-
Illinois	11/13/97	y	-	11/25/97	y	-
Indiana	9/23/97	y	-	11/25/97	y	-
Iowa	5/20/97	6/20/97	4	9/22/97	y	-
Kansas	11/05/97	11/06/97	1 day	12/17/97	1/15/98	5
Kentucky	11/20/97	y	-	2/18/97	y	-
Louisiana	11/05/97	y	-	2/19/98	y	-
Maine	5/20/97	7/28/97	10	9/22/97	11/21/97	9
Maryland	5/20/97	6/20/97	4	11/25/97	12/12/97	3
Massachusetts	12/15/97	y	-	12/17/97	y	-
Michigan	1/21/98	1/22/98	1 day	3/12/98	y	-
Minnesota	1/12/98	y	-	4/02/98	y	-
Mississippi	5/20/97	8/11/97	12	9/22/97	2/04/98	21
Missouri	7/08/97	y	-	12/17/97	y	-
Montana	5/20/97	6/20/97	4	9/22/97	y	-
Nebraska	5/20/97	6/09/97	3	9/22/97	10/15/97	5
Nevada	7/08/97	7/28/97	3	11/25/97	12/12/97	3
New Hampshire	5/20/97	9/15/97	17	11/25/97	y	-
New Jersey**	-	-	-	-	-	-
New Mexico	5/20/97	7/17/97	9	9/22/97	10/29/97	6
New York	11/18/97	11/24/97	1	1/28/98	3/24/98	8
North Carolina	7/03/97	7/31/97	4	9/22/97	y	-
North Dakota	5/20/97	7/07/97	6	9/22/97	11/24/97	10
Ohio	7/08/97	y	-	12/17/97	y	-
Oklahoma	11/18/97	12/08/97	3	1/29/98	y	-
Oregon	7/09/97	y	-	3/23/98	y	-
Pennsylvania	5/20/97	6/10/97	3	9/22/97	10/10/97	3
Rhode Island	10/16/97	10/31/97	2	10/23/97	10/31/97	1
South Carolina	5/20/97	6/20/97	4	9/22/97	10/15/97	4
South Dakota	5/20/97	6/09/97	3	9/22/97	10/06/97	2
Tennessee	10/30/97	11/06/97	1	2/19/98	y	-
Texas	6/16/97	7/17/97	4	12/17/97	y	-
Utah	5/20/97	6/20/97	4	9/22/97	12/03/97	11
Vermont	11/05/97	12/22/97	6	12/23/97	y	-
Virginia	6/16/97	7/29/97	6	12/17/97	y	-
Washington	11/19/97	y	-	3/23/98	y	-
West Virginia	6/16/97	7/02/97	2	9/22/97	10/21/97	4
Wisconsin	10/16/97	y	-	10/28/97	y	-
Wyoming	5/20/97	8/12/97	12	11/07/97	y	-
Dept. of Defense	12/17/97	y	-	12/17/97	y	-
American Samoa	7/08/97	y	-	12/17/97	y	-
Guam	5/20/97	7/01/97	6	11/24/97	y	-
North. Marianas	10/29/97	12/18/97	7	12/17/97	y	-
Puerto Rico	7/30/97	7/31/97	1 day	10/15/97	2/10/98	17
Virgin Islands	5/20/97	7/02/97	2	10/15/97	10/24/97	2

Notes: * All time intervals measured in weeks, unless specified otherwise.
**New Jersey did not submit any data files for 1996-97.
y = no response received as of 30 April 1998.

Table 2-4. Relevant Findings and Cycle Timings for 1996-97 Nonfiscal CCD Surveys

This table highlights several important findings for the 1996-97 CCD processing cycle. The total time component represents the time from initial mailout of data files to final closeout (completion of post-edit stage) of state CCD data files.

State	Use Edit System Software	Method of initial data file submission*	Lateness of initial data file submission*	Respondent Response Time (in weeks)**	Initial Mailout	Closeout (Post-edit Complete)	Total Time (in weeks)
All states, avg	-		14	22	-	-	50
Alabama	yes	I, D	30	33	3/12/97	2/03/98	48
Alaska	yes	I	18	-	2/23/97	2/03/98	50
Arizona	yes	D	20	-	2/23/97	4/02/98	58
Arkansas	yes	D, S	28	35	4/02/97	2/18/98	46
California	yes	D, S	14	-	4/02/97	3/19/98	50
Colorado	no	I	14	-	6/05/97	3/31/98	43
Connecticut	yes	I, D	28	-	2/23/97	2/12/98	51
Delaware	no	S, D	4	19	2/23/97	2/03/98	50
Dist. of Columbia	no	D	20	-	2/23/97	2/03/98	50
Florida	yes	D	3	-	2/23/97	2/03/98	50
Georgia	no	I	11	-	2/23/97	2/04/98	50
Hawaii	no	S, D	32	-	2/23/97	2/03/98	50
Idaho	no	D, S	0	-	2/23/97	2/03/98	50
Illinois	no	D	29	-	2/23/97	2/03/98	50
Indiana	yes	I, S	15	-	3/12/97	2/03/98	48
Iowa	yes	I, D	2	-	2/23/97	2/04/98	50
Kansas	yes	I, S	32	38	3/12/97	2/06/98	50
Kentucky	no	I	31	-	2/23/97	2/18/98	52
Louisiana	yes	I, D	22	-	2/23/97	2/19/98	52
Maine	yes	D	2	23	2/23/97	2/04/98	50
Maryland	yes	D	6	16	2/23/97	2/04/98	50
Massachusetts	no	D, S	20	-	2/23/97	2/04/98	50
Michigan	no	I	33	-	7/14/97	3/11/98	34
Minnesota	yes	D, S	33	-	4/02/97	4/02/98	52
Mississippi	no	D	6	42	2/23/97	2/06/98	50
Missouri	yes	I, D	3	-	2/23/97	2/04/98	50
Montana	yes	I, S	5	-	2/23/97	2/04/98	50
Nebraska	no	D	1	12	2/23/97	2/04/98	50
Nevada	yes	I	4	14	2/23/97	2/04/98	50
New Hampshire	yes	D	0	-	2/23/97	2/04/98	50
New Jersey	no	-	-	-	4/03/97	-	-
New Mexico	no	I, D	3	21	2/23/97	2/04/98	50
New York	no	D	27	34	4/03/97	3/10/98	49
North Carolina	yes	I, D	10	-	2/23/97	2/04/98	50
North Dakota	yes	D	1	20	2/23/97	2/04/98	50
Ohio	yes	I	15	-	3/12/97	2/04/98	48
Oklahoma	no	D, S	29	-	4/03/97	2/19/98	46
Oregon	yes	I, D	8	-	2/23/97	3/19/98	56
Pennsylvania	yes	D	6	15	2/23/97	2/04/98	50
Rhode Island	no	D	11	17	2/23/97	2/19/98	52
South Carolina	yes	I	0	11	2/23/97	2/13/98	51
South Dakota	yes	I	1	9	2/23/97	2/04/98	50
Tennessee	no	I	27	-	2/23/97	3/10/98	55
Texas	yes	D, S	7	-	2/23/97	2/04/98	50
Utah	no	I, S	2	19	2/23/97	2/04/98	50
Vermont	no	I, D	33	-	2/23/97	2/04/98	50
Virginia	no	I, S	12	-	2/23/97	2/04/98	50
Washington	no	I, D	7	-	4/03/97	3/19/98	50
West Virginia	yes	I	10	19	2/23/97	2/04/98	50
Wisconsin	yes	I	20	-	2/23/97	2/04/98	50
Wyoming	no	I, S	3	-	2/23/97	2/04/98	50
Dept. of Defense	no	I	32	-	2/23/97	2/04/98	50
American Samoa	no	S	1	-	2/23/97	2/04/98	50
Guam	no	I	0	-	2/23/97	2/04/98	50
North. Marianas	yes	D, S	28	-	2/23/97	2/04/98	50
Puerto Rico	no	D	15	29	4/03/97	2/10/98	45
Virgin Islands	yes	D	1	8	2/23/97	2/04/98	50

Notes: *I = internet, D = diskette, S = shuttle

*Indicates the time (in weeks) beyond the "due date" of 15 March 1997 that receipt of both the agency and school universe data files by NCES/agent occurred. This is necessary for processing to commence.

**Includes the time elapsed from mailout to receipt of the initial data files, the data summary and the edit reports. That is, the total respondent response time as depicted in Tables 2-1 and 2-3.

Table 2-5. Data Summary and Edit Phase Processing Time Interval for 1996-97 Nonfiscal CCD Surveys

This table shows the interval between receipt of initial data file from state CCD coordinators and mailout for the data summary and edit phases of the CCD processing cycle. That is, it shows the internal processing time taken by NCES and/or its agent.

State	Latest File (Agn or Sch) Received	Data Summary Mailed	Processing Interval (in weeks)	State Nonfiscal Received	Edit Mailed	Processing Interval* (in weeks)
All states, avg.	-	-	9	-	-	23
Alabama	10/08/97	10/17/97	1	10/08/97	10/24/97	2
Alaska	7/17/97	11/18/97	19	7/17/97	12/17/97	23
Arizona	7/30/97	8/06/97	2	7/30/97	4/02/98	35
Arkansas	9/30/97	10/22/97	4	9/30/97	1/27/98	18
California	6/20/97	7/08/97	3	7/23/97	3/24/98	31
Colorado	6/23/97	2/24/98	36	6/24/97	4/02/98	40
Connecticut	9/30/97	11/04/97	6	10/24/97	12/17/97	8
Delaware	4/09/97	7/08/97	14	4/07/97	11/25/97	33
Dis. of Columbia	7/30/97	11/20/97	17	7/30/97	12/17/97	21
Florida	4/07/97	5/20/97	6	4/07/97	1/08/98	39
Georgia	5/30/97	7/03/97	5	5/30/97	9/22/97	17
Hawaii	10/23/97	11/19/97	4	9/18/97	12/17/97	13
Idaho	3/10/97	5/20/97	11	8/04/97	9/22/97	7
Illinois	10/02/97	11/13/97	6	11/19/97	11/25/97	1
Indiana	6/25/97	9/23/97	13	10/30/97	11/25/97	4
Iowa	3/31/97	5/20/97	7	3/31/97	9/22/97	25
Kansas	10/24/97	11/05/97	2	10/23/97	12/17/97	8
Kentucky	10/20/97	11/20/97	5	8/27/97	2/18/98	25
Louisiana	8/14/97	11/05/97	12	10/30/97	2/19/98	16
Maine	3/25/97	5/20/97	8	3/25/97	9/22/97	26
Maryland	4/23/97	5/20/97	4	4/23/97	11/25/97	31
Massachusetts	7/31/97	12/15/97	19	10/29/97	12/17/97	7
Michigan	11/01/97	1/21/98	12	11/13/97	3/12/98	17
Minnesota	10/31/97	1/12/98	10	7/30/97	4/02/98	35
Mississippi	4/28/97	5/20/97	4	4/28/97	9/22/97	22
Missouri	4/08/97	7/08/97	13	4/16/97	12/17/97	35
Montana	4/18/97	5/20/97	4	8/04/97	9/22/97	7
Nebraska	3/24/97	5/20/97	8	3/24/97	9/22/97	26
Nevada	4/15/97	7/08/97	12	4/15/97	11/25/97	32
New Hampshire	3/18/97	5/20/97	10	3/18/97	11/25/97	27
New Jersey**	-	-	-	-	-	-
New Mexico	4/07/97	5/20/97	7	4/23/97	9/22/97	22
New York	9/23/97	11/18/97	9	9/23/97	1/28/98	19
North Carolina	5/19/97	7/03/97	7	5/20/97	9/22/97	19
North Dakota	3/25/97	5/20/97	9	3/25/97	9/22/97	26
Ohio	6/24/97	7/08/97	2	8/12/97	12/17/97	18
Oklahoma	10/01/97	11/18/97	7	10/28/97	1/29/98	13
Oregon	5/12/97	7/09/97	9	3/06/97	3/23/98	54
Pennsylvania	4/23/97	5/20/97	4	4/23/97	9/22/97	23
Rhode Island	6/03/97	10/16/97	20	6/03/97	10/23/97	21
South Carolina	3/14/97	5/20/97	9	3/14/97	9/22/97	27
South Dakota	3/19/97	5/20/97	10	3/19/97	9/22/97	28
Tennessee	9/19/97	10/30/97	6	9/24/97	2/19/98	21
Texas	4/30/97	6/16/97	6	4/30/97	12/17/97	32
Utah	3/26/97	5/20/97	8	4/09/97	9/22/97	24
Vermont	10/31/97	11/05/97	1	9/04/97	12/23/97	16
Virginia	6/05/97	6/16/97	2	10/27/97	12/17/97	7
Washington	5/05/97	11/19/97	29	5/19/97	3/23/98	44
West Virginia	5/21/97	6/16/97	4	5/21/97	9/22/97	17
Wisconsin	8/04/97	10/16/97	11	8/04/97	10/28/97	13
Wyoming	4/01/97	5/20/97	7	6/10/97	11/07/97	21
Dept. of Defense*	10/24/97	12/17/97	8	-	12/17/97	-
American Samoa	3/21/97	7/08/97	16	3/21/97	12/17/97	39
Guam	3/18/97	5/20/97	9	3/18/97	11/24/97	36
North. Marianas	9/24/97	10/29/97	5	3/18/97	12/17/97	39
Puerto Rico	6/23/97	7/30/97	6	6/23/97	10/15/97	16
Virgin Islands	3/24/97	5/20/97	8	3/24/97	10/15/97	29

Notes: * The mailing of the edit report is not contingent upon receipt of data summary report, however, the state nonfiscal data file must be received in order to generate the cross-file consistency report which is mailed with the edit report. Thus, this processing interval begins with receipt of the state nonfiscal data file. The two intervals depicted here are not cumulative.
** New Jersey did not submit any data files for 1996-97.
* DoD did not submit a state nonfiscal data file.

CHAPTER 3. AGENCY UNIVERSE AND SCHOOL UNIVERSE CCD SURVEYS PROCESSING AND EDITING STAGES

Section 3.0 Introduction

The processing and editing cycle for the evaluated surveys - the 1996-97 CCD Agency and School Universe Surveys - was characterized by five stages or phases. These stages, which to a large extent correspond to the SAS-based software programs and generated reports used to review and edit CCD survey data, included: data summary, ID assignment, internal report, main edit, and post-edit. These five stages were not necessarily mutually exclusive or sequential. Although for the purposes of this evaluation each of the stages is described individually, there was a certain degree of overlap in content as well as in the chronology of the processing and editing which took place in these stages.

The following five sections of this chapter provide a detailed evaluation of the scope, intent, associated reports, and general effectiveness of each of these five stages. Next, the data correction and response rates associated with these stages in the survey cycle are reviewed, followed by a final section of recommendations related directly to these stages.

Section 3.1 Data Summary

Upon receipt of both the agency data file and the school data file, the first phase of the 1996-97 nonfiscal CCD surveys data processing and editing cycle by NCES/agent began. In this stage, the "data summary" stage, all states' initial data submissions were converted into a form that could be processed by the SAS-based processing and editing software programs. Essentially, the data summary programs generated two reports - the "data summary report" and the "match report" - that were sent to each state CCD coordinator for his/her review and response.

The data summary report was designed to allow respondents the opportunity to view (and review) the agency and school data initially submitted by their state CCD coordinators from a slightly different perspective. That is, each record's (agency or school) identifying information and basic statistics were displayed at the aggregate-level for each particular state. This report provided information regarding any changes associated with pre-inserted data items[3] for the current year as well as prior year survey data to allow for comparison with current year data. The data summary apparently was used to detect two types of errors:

- initial data submissions which had improperly formatted data or other significant formatting errors, and
- data submissions which had significant discrepancies (at a summary level) when current and prior year data values are compared.

The match report, similar to the data summary report, was intended to serve as an edit-function for both the agency and the school data files. The match report identified missing data, potential errors, or any apparent discrepancy in the submitted data files.

This report consisted of two parts - one part each for the agency and school files. The first part indicated that verification of certain data was requested (for example, verifying that an agency operational the previous year and reported as closed for the current year was indeed closed). Second, and perhaps more importantly, the match report identified situations where resolution of a data inconsistency or "error" was required by the state. For example, a school which was listed as operational for the previous year yet was not reported for the current year, or a reported school was not associated (by identification number) with any education agency on the agency data file. These data errors or failures which required resolution were essentially "impossibilities" that needed to be corrected or resolved as quickly as possible.

The match report was critical to establishing the survey universe. If the issues identified on the match report went unresolved they would (and did) contribute to the delay of the entire current survey cycle in addition to preventing the commencement of the next survey cycle.

As indicated earlier, the states were asked to respond within two weeks with their comments and corrections

---

[3] *Pre-inserted items, for the school file, include: NCES education agency ID number, state education ID number, education agency name, state school ID number, address, city, state, zip code, phone number, school type code, operational status code, and locale code; and, for the agency file: state education agency ID number, agency name, address, city, state, zip code, phone number, education agency type code, supervisory union code, FIPS county code, CMSA/PMSA/MSA code, metropolitan status code, and boundary code.*

to the data summary and match reports that were mailed to them by NCES or its agent. Data summary "corrections" were basically adjustments to aggregate-level, basic statistics based on respondent review and comparison of current year to prior year data. There was no actual identification or indication of data "errors" that needed to be addressed by the respondent. In fact, the states were apparently not given many detailed, or specific, instructions as to what they were supposed to be scrutinizing on the data summary report, although a general cover letter requesting the states to review and respond did accompany the reports.

The match report data corrections depicted in Table 3-1 indicate the number of error messages or data failures generated by the data summary program that were considered critical and required resolution. Thus, corrections to these errors or inconsistencies were either made by the respondent or NCES collection agent regardless if a formal data summary response was submitted. Table 3-1 shows the number of all corrections to the generated reports made, by either the states or NCES/agent, at various stages in the processing and editing cycle.

Section 3.2 ID Assignment

The primary intent of this stage was to establish the survey universe for the current year survey by creating and assigning new NCES agency and school identification numbers for all newly reported schools and agencies. As with the data summary phase, both the agency and school data files must have been received in order to process the data and to generate these new identification numbers.

If the match report in the data summary stage did not indicate any problems (i.e., if there were not any data errors or failures which required resolution) with the submitted data, then the states' data files could be processed through the identification (ID) assignment program right away. Otherwise, further investigation was necessary to evaluate data inconsistencies and problems. In some cases, the data summary (including the match report) response from the state resolved the problems and allowed the ID assignment to be completed.

The SAS-based software program associated with this stage, in addition to assigning new identification numbers, also processed the agency and school files for duplicate records (duplicate identification numbers), matched schools to an education agency by state identification number, and checked for inconsistencies between data elements. The type of data problems or inconsistencies identified at this point included situations where, for example, the NCES identification number and agency boundary code did not correspond, or, where a school's operational status did not correspond with its state identification number.

The generated report was not sent to the states for review. Often the questions identified at this stage were resolved by NCES/agent staff through review of previous year data files, or via telephone, e-mail, or fax communication with respondent who provided the necessary clarifications and corrections.

Due to the fact, however, that the NCES/agent staff were required to either wait for a formal data summary response from the state or contact the state CCD coordinator directly in seeking resolution to identified data discrepancies, there were often delays in the processing and editing cycle during this stage. Again, the issues and problems detected at this point in the cycle required resolution in order to establish the survey universe and to proceed with the CCD processing and editing.

Section 3.3 Internal Report

The main objective of this stage of the processing and editing cycle for the CCD agency and school surveys was to check the submitted state data files for invalid data in preparing them for the main edit programs. That is, essentially to "clean up" the data files by examining the data in critical fields for inaccuracies and omissions. The report generated here was for internal purposes only and was not sent to the states for review.

The internal report program processed the data files by examining the data elements associated with identifying administrative information, such as name, address, identification number, etc. This program also examined the basic statistical responses, such as student and staff counts, for data fields that contained "M's," "N's," "0's," and blanks. While some of these responses were appropriate, this validation check was designed to identify any invalid response. Responses with these values in data fields, as defined in survey reporting materials, were intended to indicate the following:

M - when the data were missing and where a value is expected, but no value was measured

N - when the data were not applicable and where a value was neither expected nor measured

0 (zero) - when a numerical value was measured and no quantity was found

Blank - when data field contained no response

The internal report program also performed validity checks on the data fields in the submitted agency and school data files for other invalid nonnumeric values or out-of-range responses. For example, several of the data elements (variables) such as school file agency name (LEANM), school name (SCHNAM), agency file agency name (NAME), street name (STREET), city name (CITY) generated an error message if its length was less than or equal to three characters. Three other variables, ZIP, ZIP4, and PHONE, similarly generated error messages if the responses were less than a certain number of characters.

A review of the internal report for all 56 (New Jersey did not submit data files) states revealed some noteworthy findings. For the agency data files, 19 states generated no error messages associated with this report. For the school data files, in only five states was this the case. The error message associated with the locale code variable, indicating either a blank or inappropriate response, in the school data files was the most prevalent of all the edit messages generated. This message appeared 4,048 times in all for the 49 states' data which generated it.

The edit messages associated with a variable name having three or fewer characters, such as school or agency name, city name, or street name, was reported some 304 times for the agency and school data[4]. Interestingly, manual review of these reports indicated that all of these three-letter variable names were valid, or at least, they were not apparently changed or corrected.

As in the ID assignment stage, any error message or problem identified at this point that required resolution was either resolved by NCES/agent through a review of previous year data or referred directly to the respondents for their input. Many of the edits associated with the identifying and administrative information in this stage, while important for record completeness and accuracy, essentially performed validity checks on data which could have been checked earlier in the process. In general, however, the evaluation found this phase of the process and editing cycle to be thorough and useful, particularly in preparing the state data files for the more stringent edit checks in the next stage.

Section 3.4 Main Edit

This is the stage of the processing and editing cycle in which a comprehensive review and edit took place on initially submitted data files. These files, to varying degrees, had been reviewed, corrected, and edited via the three previously described stages in the nonfiscal CCD survey processing and editing cycle. The main SAS-based software edit program, hereafter referred to as the "main edit program", applied a series of stringent edit checks to both the agency and school data files. These edits are discussed in detail in the next chapter.

The main edit program generated three type of reports. The first, "Summary of Missing/Erroneous Data", displayed, separately for the agency and school universes, the total number of submitted records, the number of such records in error, the total number of errors, as well as the total number of errors associated with an individual edit as generated by the submitted data files. Although this report was not sent to the states for their review, scrutiny of these edits constituted a large portion of this evaluation and will be covered in detail in the next chapter.

The "Cross-file Consistency Report", the second generated report, provided a comparison of various corresponding data category totals between the reported state nonfiscal data and both the agency and school totals. This report, and the 1996-97 State Nonfiscal CCD Survey, are addressed in Chapter 6.

The third report, the "Edit Report", provided a listing of edit messages associated with identified errors for the top 50 most error-filled agency records, and the 100 most error-filled school records for each state. Errors listed in the edit report were identified with an edit status of "C", indicating a critical error, or "W" for a warning error. The weighting system, which determined the highest weighted in-error records, was predicated on the assignment to a critical error a value of "10", and a warning message a value of "1".

These edit reports, along with the cross-file consistency report, were sent to the states for their review. Respondents were asked to make manual corrections and to provide their comments within two weeks. Table 3-1 displays the number of respondent-made corrections as a result of this review. The non-response issue will be discussed later in this chapter.

[4] *Of the 304 total displayed error messages, 238 were associated with city name, 36 with street name, and 30 with agency or school name.*

Section 3.5 Post-Edit

The stated purpose of this final phase of the processing and editing cycle was, as an edit-function, to determine if any additional "critical" data changes or corrections were necessary. This phase was also used to ensure that all response fields in the survey contained a response before the files were merged for delivery to NCES as complete, final data files (these data files were also included in the following survey year's initial mailout as "prior year" data). The post-edit program essentially performed the following edit functions:

- Converted remaining blanks in pre-inserted items to prior year data (if there was a match between current and prior year) or to "M".
- Converted all closed school data to "N".
- Converted remaining "blanks" (no response provided) to "M".
- Where there were data in some categories (like student total membership by grade provided for some but not all grades), converted "N" to zero (0).
- Where there were no data in any category, converted all "M's" or zeroes (0's) to "N".

The post-edit program also generated several reports, including some 28 summary reports, which provided details concerning the changes and conversions (e.g., the number of records where classroom teachers was blank and converted to "M") made to each state's data files at this point. Some reports provided information, or identification, of remaining data inconsistencies, such as schools records not associated with any agency, a blank or missing metropolitan status code, or a current year FIPS code that did not match the previous year, that required additional research and resolution (a manual review of these post-edit summary reports was on-going at the time of this evaluation).

All of these reports were intended for internal review and were not sent to the states for comment or review. In a few instances, these reports appeared to identify issues which should have been resolved earlier in the processing and editing cycle. In any event, for the observed processing and editing cycle, apparently due to time and resource constraints, very little use was made of these reports at this time. Later, the post-edited data were presented in several tables to NCES for their review and verification.

A second data summary report was generated, in part, to prepare the data file to be processed through the post-edit program. A final cross-file consistency report and internal report, as well as a third data summary report (the same format and presentation as the previous two) were also generated by the post-edit program. This evaluation observed these reports not to be used for any tangible evaluative purpose for this survey cycle.

The apparent purpose (and use) of the post-edit for the observed cycle was to prepare the states' data files for the subsequent year mailout, in this case for the 1997-98 Nonfiscal CCD Surveys. Earlier delays and the general "lateness" of the current survey, particularly as it affected the subsequent survey year's mailout, caused the processing for most states at this point to be rather cursory. The primary focus at this stage in the survey cycle apparently was not one of edit, but simply to complete as final all remaining states' data files.

This evaluation did not address any edits in use beyond the post-edit phase. Thus, any editing done by NCES to assist in the review of tabulated data, such as aggregates for publication in printed reports, was beyond the scope of this evaluation. Apparently, however, these processes do add considerable time to the overall cycle.

Section 3.6 1996-97 CCD Data Correction and Response Rates

One of the more remarkable aspects of this survey cycle was the very high non-response rates exhibited by the participants. This was particulary in evidence for the data summary phase and, even more so, for the edit phase. In both instances, the states were requested to respond by reviewing and correcting reports which had been mailed to them for their comments.

Type of submission	Non-respondents	
	No.	%
Initial data files	1	1.8
Data summary report response	17	29.8
Edit report response	37	64.9

For the 1996-97 Nonfiscal CCD Surveys cycle while only one state, New Jersey, did not submit initial data files, almost one-third did not respond to the data summary response request. About two-thirds of the states failed to provide an edit response as shown in the above chart. These response rates were based on verifiable documented (marked-up originals or fax transmissions in most cases) responses returned by the respondents. It was likely that some data issues or questions were handled by informal (undocumented) communication between the respondent and NCES or its collection agent. The implications of such low participation rates in these critical editing procedures clearly speak to the issues of survey completeness and accuracy.

Given the low response rates, the very small number of changes or corrections attributed to respondent review was not found to be all that surprising. Indeed, the 33 total corrections made at data summary phase and the 219 corrections at the edit phase, as presented in Table 3-1 for both the agency and school files, represented a very minute number and percentage, especially when considering the volume of errors that were generated and potentially reviewed. This was especially true for the edit report corrections[5].

Tables 3-2 and 3-3 present the number of total records, the number of in-error records and the percentage of these records ultimately reviewed and corrected at the edit stage for both the agency and school surveys. Some rather notable figures stand out.

For the agency universe, of the submitted 15,848 records from the 56 states that provided initial data files, 9,891, or just over 60 percent, contained at least one error as generated by the CCD main edit program. The 20 states which responded to the edit reports actually reviewed slightly less than 40 percent of the error-filled records, while correcting only about two-tenths of one percent (54 corrections divided by the 20,591 total errors for responding states) of the total errors. These percentages dropped to 20 percent and a very minuscule 0.3 percent when all states (including the non-respondent states) were taken into account.

For the school universe, of the 88,987 records submitted, 27,387 of them contained errors for an in-error rate of about 30 percent. Similar to the agency file figures, about 46 percent of error-filled records were reviewed and 3 percent of total errors corrected for the 20 responding states, and 18 percent and 0.3 percent for all states, for the school survey data in Table 3-3.

Important to remember, the states were not mailed all of their records that contained errors - only the top 50 error-filled agency records (top 100 for school record files) were returned for comment. Due to this fact, less than one-half of the states were provided the opportunity to review all of their agency records which generated errors during the edit stage processing. Only about 20 percent of the respondents had the opportunity to review all of their error-filled school records. Several of the larger states, including California, Illinois, Michigan, and Texas among others, had the opportunity to review less than 10 percent of their agency and/or school in-error records. Table 3-3 provides a state-by-state summary of these percentages.

NCES/agent survey data processing staff indicated that the low return and correction rates, particularly for the data summary and edit report requests, stemmed from unique situations in various states which resulted in the same errors being generated year after year. In short, it was suggested that this occurrence indicated the respondents were, generally speaking, at odds with having to deal with the same non-issues every year[6]. Regardless, the fact that less than 20 percent of all in-error data records were ever reviewed by the respondents and that about one-half of all the materials mailed to the states after their initial data submissions were not returned indicated serious potential shortcomings in data quality and overall accuracy of the surveys.

[5] *The cross-file consistency related correction rates will be discussed in Chapter 6.*

[6] *A technical review panel conducted in November 1997 by NCES on the topic of editing and processing with several state CCD coordinators generally supported this notion as well as several of the recommendations made in this chapter.*

Section 3.7 Recommendations

Review of the five primary stages which made up the nonfiscal CCD survey cycle resulted in several recommendations. It is suggested that any changes or revisions to the processing and editing activities which occur as a result of the anticipated CCD survey expansion for 1998-99 should include consideration of the following:

1. Given the very low survey follow up response and error-correction rates along with the condition that many error corrections were apparently initiated and addressed by NCES/agent, the efficacy of the existing post-submission edit and review process is brought into question. A radical approach would involve revamping the edit process to eliminate (or significantly reduce), as much as possible, the procedure of returning edit materials/reports back to the states for their review. Such action would more formally place the burden and responsibility of data editing on NCES and its agent. This is to a large extent, however, what occurs anyway.

2. Other, less radical, recommendations would include adjustments to the current procedures. Several of these recommendations relate directly to the physical layout/presentation of the current data summary report. Alternative format suggestions include:

    a. Provide more detailed descriptions of what is expected of the respondent in terms of reviewing and correcting the data presented in this report, including highlighting the due date, directly on the report.

    b. Delete the "five smallest and largest values" headings and data for all data categories from the report.

    c. Place current year data and prior year data side-by-side and arrange dropout data totals in a tabular form (less the five smallest/largest data categories) to enhance ease of data comparison.

    d. For the school universe report, add a heading - "OPERATIONAL STATUS" - above the listing of schools (new, closed, etc.).

    e. For the agency universe report, add a heading - "BOUNDARY STATUS" - above the listed LEA data categories.

3. More clearly highlight the allotted two week response time interval requested for state response, perhaps specifying an actual due date, in the instructions at the top of the match report.

4. Add an edit into the Edit System software program to compare assigned state and NCES identification numbers. This would require that the database listing of previously assigned identification (ID) numbers be provided with the initial reporting materials, but would eliminate the identification conflict "impossibilities" (such as a state education agency ID number being associated with more than one NCES education agency ID number on the school file) that appeared on the match report, often requiring resolution by the respondent and resulting in time delays.

5. Provide respondents with capability, when necessary, to assign their own new ID numbers. This could, among other possibilities, be accomplished by:

    a. Modifying Edit System software to generate new NCES agency and school ID numbers from the ALLLEAID and ALLSCHNO data files (the files from which new ID numbers are assigned) as states add new agencies or schools to their survey universes.

    b. Providing states access to the ALLLEAID and ALLSCHNO data files directly via the internet.

6. Perform the internal report "edit" earlier in the editing and processing cycle. This would provide a barometer of the quality of a state's submission. Any potential data issues might be addressed sooner, perhaps included as part of the data summary program and report mailout, thereby eliminating the current practice of having to repeatedly contact respondents directly.

7. Modify the validity edits in the internal report program (similar to all subsequent edit programs) that examine identifying information data, like agency or school name, city name, or street name. It is recommended that the defining criteria be changed to identify responses having two characters or less rather than three characters or less. Although these certainly are not critical edits, this change would eliminate many of the error messages that state and federal reviewers and analysts must filter.

8. Replace the existing edit reports (which consist of the top 50 agency and top 100 school error- filled records) with an edit "summary report." It is recommended that such a report indicate the record (agency or school) ID number, the data response in question, the error type (critical or warning), and the error message generated for all in-error records. This action would accomplish two important feats:

    a. Reduce the volume (and length) of edit materials requiring review that would be sent back out to the states, thus reducing respondent burden and likely improving response rates.

    b. Ensure that all in-error records (and individual errors) would be available for respondent review.

9. In order to make more productive use of the post-edit stage, consideration should be given to thoroughly reexamining the 28 generated post-edit summary reports for their intended purpose, function, and usefulness, as a component of the processing and editing cycle.

Table 3-1. Summary of Number of Corrections Made at Data Summary and Edit Phases to Submitted Data Files in 1996-97 Nonfiscal CCD Surveys

This table shows the occurrence and number of data corrections made by respondent and NCES collection agent, based on state data submissions, at various stages in the 1996-97 Nonfiscal CCD Surveys data processing and editing cycle.

State	Pre-Edit Reports						Edit Reports			
	Data Summary		Match Report*		Internal Report		Cross-File		Edit Report	
	Agency	School	Agency	School	Agency	School	St- Agn	St.-Sch	Agency	School
Total	12	21	372	1,869	35	97	8	5	54	165
Alabama	1	1	0	0	0	0	1	0	0	0
Alaska	2	0	0	7	1	0	na	na	na	na
Arizona	na	na	79	26	0	0	na	na	na	na
Arkansas	1	0	17	3	0	0	0	0	16	103
California	0	0	3	26	0	0	na	na	na	na
Colorado	0	0	3	126	1	2	na	na	na	na
Connecticut	0	0	1	5	0	57	na	na	na	na
Delaware	0	7	0	1	1	1	1	0	1	3
Dis. of Columbia	na	na	0	2	0	0	na	na	na	na
Florida	0	0	0	100	0	0	na	na	na	na
Georgia	0	0	3	29	0	0	na	na	na	na
Hawaii	na	na	0	0	0	0	na	na	na	na
Idaho	0	0	0	10	0	0	na	na	na	na
Illinois	na	na	0	3	0	0	na	na	na	na
Indiana	na	na	0	13	0	0	na	na	na	na
Iowa	0	0	0	11	0	1	na	na	na	na
Kansas	0	0	0	46	0	0	0	0	0	0
Kentucky	na	na	79	0	0	0	na	na	na	na
Louisiana	na	na	0	25	0	0	na	na	na	na
Maine	0	0	0	0	0	0	0	4	0	1
Maryland	0	0	0	12	0	1	0	1	0	0
Massachusetts	na	na	1	0	0	0	na	na	na	na
Michigan	0	0	37	587	0	0	na	na	na	na
Minnesota	na	na	61	277	12	0	na	na	na	na
Mississippi	0	0	0	3	0	0	0	0	3	5
Missouri	na	na	5	0	0	0	na	na	na	na
Montana	0	0	0	0	0	0	na	na	na	na
Nebraska	0	0	1	3	0	0	1	0	12	0
Nevada	0	0	0	2	0	7	1	0	0	2
New Hampshire	0	0	3	4	3	1	na	na	na	na
New Jersey**	-	-	-	-	-	-	-	-	-	-
New Mexico	0	0	0	1	0	1	3	0	9	2
New York	1	1	3	10	0	1	1	0	2	0
North Carolina	0	0	0	11	0	0	na	na	na	na
North Dakota	0	0	0	3	0	0	0	0	0	0
Ohio	na	na	5	20	0	0	na	na	na	na
Oklahoma	0	0	3	5	0	0	na	na	na	na
Oregon	na	na	32	113	0	0	na	na	na	na
Pennsylvania	2	0	1	6	1	0	0	0	0	0
Rhode Island	0	1	0	0	0	0	0	0	0	1
South Carolina	0	0	0	27	0	0	0	0	9	3
South Dakota	0	0	1	4	0	0	0	0	0	0
Tennessee	0	5	0	160	0	0	na	na	na	na
Texas	0	0	17	32	0	0	na	na	na	na
Utah	0	0	0	0	0	1	0	0	0	4
Vermont	0	0	12	9	5	0	na	na	na	na
Virginia	2	0	5	0	4	0	na	na	na	na
Washington	na	na	0	92	7	0	na	na	na	na
West Virginia	0	0	0	0	0	6	0	0	0	0
Wisconsin	na	na	0	0	0	0	na	na	na	na
Wyoming	0	0	0	1	0	0	na	na	na	na
Dept. of Defense	na	na	0	11	0	0	na	na	na	na
American Samoa	na	na	0	0	0	0	na	na	na	na
Guam	3	6	0	0	0	0	na	na	na	na
North. Marianas	0	0	0	0	0	0	na	na	na	na
Puerto Rico	0	0	0	43	0	18	0	0	2	38
Virgin Islands	0	0	0	0	0	0	0	0	0	3

Notes: *Indicates the number of error messages generated by data summary program that required resolution and, thus, were either corrected by the respondent or NCES collection agent regardless if a formal data summary response was submitted by respondent.
**New Jersey did not submit any data files for 1996-97; na = not available, no response was received as of 30 April 1998.

Table 3-2. Summary of Total Number of Records in Error and Total Number of Errors Generated from Main Edit Program for Submitted Data Files in 1996-97 Nonfiscal CCD Surveys

This table shows the number of agency and school data records in error at the edit stage in the 1996-97 Nonfiscal CCD Surveys data processing cycle. It also shows the number of error and warning messages these records generated as well as the percentage of records in error.

State	Agency Universe				School Universe			
	Number of Records	Number of Records in Error	Percent of Records in Error	Total Number of Errors	Number of Records	Number of Records in Error	Percent of Records in Error	Total Number of Errors
Total/Average*	15,848	9,891	62.4	20,591	88,987	27,387	30.8	52,228
Alabama	131	21	16.0	51	1,349	179	13.3	300
Alaska	55	24	43.6	37	505	173	34.3	267
Arizona	317	190	59.9	561	1,357	335	24.7	553
Arkansas	332	89	26.8	190	1,107	105	9.5	191
California	1,060	795	75.0	1,235	8,004	1,620	20.2	2,260
Colorado	194	50	25.8	63	1,540	323	21.0	521
Connecticut	179	50	27.9	70	1,065	128	12.0	213
Delaware	24	24	100.0	46	184	44	23.9	76
Dis. of Columbia	1	1	100.0	2	187	101	54.0	149
Florida	74	27	36.5	53	2,880	642	22.3	1,017
Georgia	183	37	20.2	39	1,831	435	23.8	766
Hawaii	1	1	100.0	1	249	42	16.9	103
Idaho	113	19	16.8	30	640	629	98.3	734
Illinois	1,046	864	82.6	1,904	4,229	596	14.1	922
Indiana	327	64	19.6	118	1,959	224	11.4	380
Iowa	415	199	48.0	318	1,564	225	14.4	321
Kansas	304	97	31.9	107	1,525	180	11.8	290
Kentucky	258	258	100.0	855	1,454	222	15.3	391
Louisiana	72	32	44.4	69	1,487	178	12.0	272
Maine	326	219	67.2	563	729	108	14.8	167
Maryland	24	9	37.5	9	1,289	104	8.1	177
Massachusetts	465	465	100.0	2,047	1,880	1,856	98.7	3,911
Michigan	730	730	100.0	1,869	4,006	947	23.6	1,733
Minnesota	459	395	86.1	703	2,159	1,062	49.2	1,869
Mississippi	164	80	48.8	108	1,027	236	23.0	360
Missouri	532	119	22.4	172	2,315	424	18.3	681
Montana	557	280	50.3	650	899	233	25.9	282
Nebraska	786	396	50.4	752	1,423	518	36.4	910
Nevada	18	7	38.9	16	444	91	20.5	168
New Hampshire	248	248	100.0	582	517	82	15.9	114
New Jersey**	-	-	-	-	-	-	-	-
New Mexico	89	24	27.0	28	734	69	9.4	100
New York	756	317	41.9	402	4,221	548	13.0	955
North Carolina	123	102	82.9	160	2,018	164	8.1	271
North Dakota	285	73	25.6	119	615	148	24.1	239
Ohio	781	767	98.2	1,616	3,922	620	15.8	1,024
Oklahoma	550	549	99.8	757	1,839	272	14.8	375
Oregon	249	62	24.9	96	1,231	177	14.4	240
Pennsylvania	615	117	19.0	205	3,213	315	9.8	585
Rhode Island	37	8	21.6	9	316	71	22.5	100
South Carolina	106	28	26.4	45	1,088	170	15.6	272
South Dakota	219	86	39.3	166	851	271	31.8	381
Tennessee	140	140	100.0	465	1,582	1,565	98.9	3,470
Texas	1,060	1,059	99.9	1,394	6,979	6,849	98.1	18,641
Utah	47	39	83.0	61	749	77	10.3	133
Vermont	346	138	39.9	407	400	151	37.8	227
Virginia	169	169	100.0	597	1,931	1,895	98.1	2,185
Washington	305	305	100.0	621	2,180	547	25.1	920
West Virginia	57	18	31.6	28	883	133	15.1	203
Wisconsin	444	65	14.6	88	2,119	237	11.2	385
Wyoming	58	18	31.0	35	413	82	19.9	109
Dept. of Defense	12	12	100.0	63	173	165	95.4	225
American Samoa	1	1	100.0	1	31	31	100.0	34
Guam	1	1	100.0	2	35	16	45.7	27
North. Marianas	1	1	100.0	1	26	13	50.0	35
Puerto Rico	1	1	100.0	2	1,599	754	47.2	983
Virgin Islands	1	1	100.0	3	35	5	14.3	11

Notes: *For number of records, number of records in error, and total number of errors these values are totals. The percentage of records in error is an average of all states and outlying areas.
**New Jersey did not submit any data files for 1996-97.

Table 3-3. Summary of Percentage of Records in Error Reviewed and Errors Corrected by Respondents During Edit Phase in 1996-97 Nonfiscal CCD Surveys

This table shows the percentage of agency and school data records in error and the percentage of records in errors reviewed and corrected by respondent during the edit stage in the 1996-97 Nonfiscal CCD Surveys data processing cycle.

State	Agency Universe				School Universe			
	No. of records in error	%of in-error records reviewed*	No. of errors corrected	% of record errors corrected	No. of records in error	%of in-error records reviewed*	No. of errors corrected	% of record errors corrected
Total/Avg.								
States responding	1,647	38.3	54	2.2	3,578	46.3	165	2.9
All states	9,891	20.3	54	0.3	27,387	18.3	165	0.3
Alabama	21	100.0	0	0.0	179	55.9	0	0.0
Alaska	24	100.0	na	-	173	57.8	na	-
Arizona	190	26.3	na	-	335	29.9	na	-
Arkansas	89	56.2	16	8.4	105	95.2	103	53.9
California	795	6.3	na	-	1,620	6.2	na	-
Colorado	50	100.0	na	-	323	31.0	na	-
Connecticut	50	100.0	na	-	128	78.1	na	-
Delaware	24	100.0	1	2.2	44	100.0	3	3.9
Dis. of Columbia	1	100.0	na	-	101	99.0	na	-
Florida	27	100.0	na	-	642	15.6	na	-
Georgia	37	100.0	na	-	435	23.0	na	-
Hawaii	1	100.0	na	-	42	100.0	na	-
Idaho	19	100.0	na	-	629	15.9	na	-
Illinois	864	5.8	na	-	596	16.8	na	-
Indiana	64	78.1	na	-	224	44.6	na	-
Iowa	199	25.1	na	-	225	44.4	na	-
Kansas	97	51.5	0	0.0	180	55.6	0	0.0
Kentucky	258	19.4	na	-	222	45.0	na	-
Louisiana	32	100.0	na	-	178	56.2	na	-
Maine	219	22.8	0	0.0	108	92.6	1	0.0
Maryland	9	100.0	0	0.0	104	96.2	0	0.0
Massachusetts	465	10.8	na	-	1,856	5.4	na	-
Michigan	730	6.8	na	-	947	10.6	na	-
Minnesota	395	12.7	na	-	1,062	9.4	na	-
Mississippi	80	62.5	3	2.8	236	42.4	5	1.4
Missouri	119	42.0	na	-	424	23.6	na	-
Montana	280	17.9	na	-	233	42.9	na	-
Nebraska	396	12.6	12	1.6	518	19.3	0	0.0
Nevada	7	100.0	0	0.0	91	100.0	2	1.2
New Hampshire	248	20.2	na	-	82	100.0	na	-
New Jersey*	-	-	-	-	-	-	-	-
New Mexico	24	100.0	9	32.1	69	100.0	2	2.0
New York	317	15.8	2	0.5	548	18.2	na	-
North Carolina	102	49.0	na	-	164	61.0	na	-
North Dakota	73	68.5	0	0.0	148	67.6	0	0.0
Ohio	767	6.5	na	-	620	16.1	na	-
Oklahoma	549	9.1	na	-	272	36.8	na	-
Oregon	62	80.6	na	-	177	56.5	na	-
Pennsylvania	117	42.7	0	0.0	315	31.7	0	0.0
Rhode Island	8	100.0	0	0.0	71	100.0	1	1.0
South Carolina	28	100.0	9	20.0	170	58.8	3	1.1
South Dakota	86	58.1	0	0.0	271	36.9	0	0.0
Tennessee	140	35.7	na	-	1,565	6.4	na	-
Texas	1,059	4.7	na	-	6,849	1.5	na	-
Utah	39	100.0	0	0.0	77	100.0	4	3.0
Vermont	138	36.2	na	-	151	66.2	na	-
Virginia	169	29.6	na	-	1,895	5.3	na	-
Washington	305	16.4	na	-	547	18.3	na	-
West Virginia	18	100.0	0	0.0	133	75.2	0	0.0
Wisconsin	65	76.9	na	-	237	42.2	na	-
Wyoming	18	100.0	na	-	82	100.0	na	-
Dept. of Defense	12	100.0	na	-	165	60.6	na	-
American Samoa	1	100.0	na	-	31	100.0	na	-
Guam	1	100.0	na	-	16	100.0	na	-
North. Marianas	1	100.0	na	-	13	100.0	na	-
Puerto Rico	1	100.0	2	100.0	754	13.3	38	3.9
Virgin Islands	1	100.0	0	0.0	5	100.0	3	27.3

Notes: *Not all records in error are reviewed by the respondents. For the agency survey only the top 50 error-filled records (by error weight), and for the school survey only the top 100 error-filled records, are sent back to the respondents for review.
*New Jersey did not submit any data files for 1996-97; na= not available, no response was received as of 30 April 1998.

CHAPTER 4. CCD MAIN EDIT PROGRAM EDITS AND RELATED ERROR AND WARNING MESSAGES

Section 4.0 Introduction

This chapter pertains to the CCD main edit program[7] edit checks and the error and warning messages they generated during the processing and editing cycle of state data submissions for the 1996-97 Nonfiscal CCD Surveys. The edit checks and associated error and warning messages described in this chapter related directly to the states' original data submissions as processed in the edit phase for the 1996-97 census.

The CCD main edit program can generate nearly 90 error/warning messages. Appendices A and B provide a detailed listing of the 54 education agency universe edits and the 34 school universe edits, the error/warning messages produced, and the condition that generated them[8]. The tabulations and error/warning messages rates presented in this chapter were the result of the CCD main edit program being applied to the states' initial data submissions. These submissions were processed and edited, to varying degrees, during preceding cycle stages described in Chapters 2 and 3.

Based on the 1996-97 CCD nonfiscal submissions, files containing 15,848 agency universe data records and 88,987 school universe data records generated a total of 20,591 and 52,228 error and warning messages, respectively. For a majority of the states (35), or 63 percent, the agency data files contained between 1 and 200 errors, while 6 states (about 11 percent) had more than 1,000 errors generated by their agency file records. Shown below, the number of errors generated by the states' school records was even greater in that 36 states (about 65 percent) had up to 400 total errors and 9 (16 percent) others had more than 1,000 errors.

Total no. of errors	Agency		School	
	No.	%	No.	%
Total*	56	100.0	56	100.0
1 to 200	35	62.5	16	28.6
201 to 400	2	3.6	20	35.7
401 to 600	7	12.5	4	7.1
601 to 800	5	8.9	2	3.6
801 to 1,000	1	1.8	5	8.9
> 1,000	6	10.7	9	16.1

*New Jersey did not submit any data files for 1996-97.

The submitted agency data files contained 15,848 records, some 5,957 (38 percent) of which generated no error messages at all. The number of error messages generated by a single agency record ranged from none to 12, with more than 55 percent having 1 to 3 error messages. Of the 88,987 school universe records submitted, 61,600, or almost 70 percent, passed every edit condition. Just slightly over 2 percent (1,927) of the records had more than 4 error messages.

While the following charts show the range of errors per record for the two universes, state-by-state totals are found in Table 3-2 at the end of Chapter 3:

Number of error messages	Agency universe records	
	No.	Percent
Total*	15,848	100.0
More than 12	0	0.0
10 to 12	22	0.1
9 to 11	312	2.0
4 to 6	745	4.7
1 to 3	8,812	55.6
No error messages	5,957	37.6

*Total based on number of records in submitted data files.

[7] *This editing software program differs slightly from the Edit System software that is provided to the respondents to assist in their initial data submission. This software is discussed in a later section of this chapter.*

[8] *The numbers assigned to the edits in this evaluation correspond to the order in which the edits appear in the main edit program and are strictly for identification and tracking purposes.*

Number of error messages	School universe records	
	No.	Percent
Total*	88,987	100.0
More than 12	0	0.0
10 to 12	6	(z)
7 to 9	15	(z)
4 to 6	1,906	2.1
1 to 3	25,460	28.6
No error messages	61,600	69.2

*Total based on number of records in submitted data files.
(z) less than .05 percent.

Section 4.1 Application of the Edits to Data Collection

Table 4-1, for the agency universe, and Table 4-2, for the school universe, list the 60 and 36 respective edit checks and the number of error/warning messages generated by each that were applicable to the 1996-97 Nonfiscal CCD Surveys[9]. These tables also indicate the type and source - administrative information, code classification, or basic agency/school statistics -of data problem detected by the edit as well as the number of states whose data files generated each particular error. It is interesting that about one-fourth of the edit checks generated errors by at least 30 states.

An important note about the evaluation of the edits based on the state data submissions is that the counts of error and warning messages in this chapter reflected the data as submitted by the state CCD coordinators. These submissions were not necessarily the raw data often associated with a data collection. In some cases for these surveys, the first level of editing was performed by state coordinators during data entry via the provided Edit System software, or their own state's preferred software. The edits described in this chapter were the ones associated with the main edit program used during the editing and processing cycle. As a result, the tallies cited in this evaluation reflected data records that were, in some instances, subject to edit at least once prior to submission. No accurate measure was available of the error warning messages that may have been generated during the initial data entry by the state CCD data coordinators. Thus, it was not possible in this evaluation to fully quantify the effectiveness of the particular edits as they pertained to the raw data - that is, the basic information compiled by state data as they assembled their respective data files for transmission to NCES or its agent.

Table 4-3 provides a summary of the error/warning messages that were generated by the main edit program when the data submitted by the states were processed by the main edit software program. In this table, the edits were grouped into three descriptive categories depicting the source of potential problems identified. The first category, "Administrative information," covered information related to identification such as agency (or school) name, state or NCES identification number, address, telephone number, etc. A second category, "Code classification information," referred to various codification-related data elements, such as agency type, boundary code, school operational status, type code, FIPS code, or metropolitan status code, and the like.

A final category, the "Basic agency (or school) statistics," was further subdivided into three groups - "student counts," "staff counts," and "agency (or school)-wide counts." This last subgroup identified such edits/data elements which did not solely, and specifically, relate to either student or staffing measures, but rather applied to the entire agency (or school). These measures included, for example, pupil/teacher ratio comparisons, no teachers and students reported, or an agency with no schools or data. This category was of particular importance because it represented the primary statistical data collected by the CCD nonfiscal surveys.

Table 4-3 indicates that, for both the agency and school universe data files, slightly more than nine out of every 10 error generated by the main edit program in the edit phase of the survey cycle was attributable to the basic statistic category. For the two surveys combined, of a total of 72,819 error messages, the category measuring the basic agency (or school) statistics generated 67,183 of these errors. There were several observations about the messages for this most critical category:

1. They represented slightly more than 92 percent of all the errors messages generated.

2. Within this category, the "student counts" group generated the largest number of error messages. These comprised 60 percent of this category's error messages for the agency files and 43 percent for the school files. They accounted for 56

[9] *For the agency survey, six edits, and, for the school survey, two edits are no longer used.*

percent and 40 percent, respectively, of all error and warning messages.

3. The 5 error messages representing the "agency (or school)-wide" subgroup generated 14,258 (20 percent) of the error messages.

The "code classification" category of edits generated less than one-half of one percent of all error messages when the agency and school data files were processed. This minuscule percent suggested that any problems identified by these edits were resolved at data entry by the state data coordinators and corrected before submission or resolved by NCES and its agent at earlier stages in the editing and processing cycle.

The "administrative information" category of edits generated the second largest percentage of the generated error messages at seven and eight percent, respectively, for the agency and school survey data. While these edits represented less than ten percent of the total errors generated, nearly one-third (27 of the 88 total edits) of all the edits were dedicated to checking administrative-type data. The extensive and thorough editing on administrative information was a strong point in the CCD edit process. As with the code classification-related edit category, the relatively small number of administrative errors would seem to indicate that, to a large extent, these data elements were being properly screened prior to data submission and/or corrected prior to the main edit at earlier stages in the editing process.

Section 4.2 CCD Main Program Data Edits by Type

For the purposes of this evaluation, the edits performed by the CCD main edit program were grouped into three types of data checks:

- Relational comparisons to other data elements, including out-of-range values
- Validation edits, often blank, missing, not applicable, or zero
- Historical comparisons to prior year data

Although to some degree these categories overlap, the edits were categorized into these groups based on their primary purpose.

In the charts below, the number of edit checks and error/warning messages represented by these three types of generated main edit program edits from the 1996-97 Agency Universe and School Universe CCD Surveys submitted data are shown:

Agency universe -

Type of edit	Edit checks		Error/warning messages	
	No.	%	No.	%
Total	54	100.0	20,591	100.0
Relational	22	40.7	9,429	45.8
Validation	26	48.1	8,758	42.5
Historical	6	11.1	2,404	11.7

School universe -

Type of edit	Edit checks		Error/warning messages	
	No.	%	No.	%
Total	34	100.0	52,228	100.0
Relational	11	32.4	31,508	60.3
Validation	21	61.8	20,508	39.3
Historical	2	5.9	212	0.4

The following three sections of this chapter examine these edit types in more detail.

Section 4.3 Relational Edits

This type of edit generally involved comparing two or more data elements. The comparison typically consisted of a calculation using the values of the reported data elements. The calculation (e.g., a ratio between two values) was expected to fall within an acceptable, predetermined range. It was usual for the acceptable ranges to be based upon empirical evidence from the current survey or from prior surveys. Edit checks of this type included, for example, a comparison of pupil/teacher ratio between the current and prior years to note any change (difference) greater than ±20 percent, or, an indication that the total students by grade changed by more than ±25 percent between the current and prior years. In both instances these occurrences prompted an "error" message.

The charts in the preceding section, categorized and displayed by survey, indicated for the combined 1996-97 agency and school surveys that relational edits

accounted for over one-third (38 percent) of the edit checks, and 56 percent of the total error and warning messages generated. Tables 4-1 and 4-2 identify the individual edits that comprised these categories.

Table 4-4 displays, for the agency survey, that while the 9,429 error messages in this category represented 46 percent of all the generated messages, there was great dispersion in number of errors among the various edits. The top three most frequently occurring edits (#59, 40, and 50) accounted for over 53 percent of this group and nearly a quarter of all generated error messages. On the other hand, 13 of the 22 edits of this type accounted for less than 1 percent of all generated errors, with 4 of them (#01, 18, 25, and 58) generating no error messages at all. Table 4-5 reveals similar findings for the school survey, with 2 edits (#20 and 36) generating 20,192 error messages, thus accounting for 64 percent of this category's 31,508 error messages. Nearly 40 percent of all the error/warning messages were generated by two edit checks: pupil/teacher current year to prior year ratio exceeding ±20 percent (12,509) and total teacher difference between current and prior year exceeded ±25 percent (7,683).

The relational edits in the CCD main edit program were found to be very thorough. Although no specific additional relational edits are suggested for inclusion to the existing main edit program, there was one area of concern. As noted in the previous paragraph and in Tables 4-4 and 4-5, a disproportionate number of error messages were generated by relational edits which contained an "acceptable" range/tolerance parameter (e.g., current year data value compared to prior year data value exceeded a ±25 percent difference). It is recommended that all edits containing such a criterion parameter be reevaluated for appropriateness and effectiveness. A school with only 15 to 20 students, for example, would be highly sensitive to percent changes. Perhaps an absolute change value or a minimum value for comparison, such as setting minimum enrollment and staffing numbers for application of edits, could be introduced. Such alterations would likely ensure more useful checks aimed at identifying true data errors and outliers while preventing excessive erroneous data failures from being generated. These recommendations are discussed in more detail in the next chapter.

Section 4.4 Validation Edits

Validation edits accounted for a majority of the edit checks in the CCD main edit program (56 percent) but only 40 percent of the error and warning messages generated by the 1996-97 nonfiscal CCD submitted data. This type of edit typically involved testing data elements to ascertain if a level of "reasonableness" existed, often checking for missing information. Theoretically, at least in most instances, education agencies should have had associated schools, schools should have had teachers, staff, students, physical location, and so on. Also, each agency and school should have been assigned unique state and NCES identification numbers. These edits essentially looked for blank or zero values for most data elements.

However, to a large extent, these edits do not examine whether the data responses are reasonable, blank and zero values notwithstanding. Rather they were primarily screening the data for unreasonably missing data. As shown in Tables 4-4 and 4-5, 8,758 and 20,508 errors messages were generated by the category of validation edits for the agency and school surveys, respectively. In the agency survey (Table 4-4), of the 26 edits which comprised this category, 17 generated error messages accounting for less than 1 percent each of the total error and warning messages. Eight of these edits actually generated no error messages at all. The top 4 edits (#12, 39, 44, and 19), as a percentage of total errors generated in this category, produced 5,416 (62 percent) of the 8,758 error messages attributed to it. Together these four edits accounted for slightly more than a quarter of all the error messages generated.

Table 4-5 shows that 5 edits (#18, 27, 34, 31, and 16) produced more than 2,000 error messages each and represented over 90 percent (18,720 errors out of 20,508) of this category's total. On the other hand, 15 of the remaining 16 edits which made up the validation checks category failed to generate sufficient numbers of error messages to account for even 1 percent of the total warnings generated. These 16 edits, together, only generated 1,788 error messages, or less than 4 percent of the total. In conclusion, a relatively small number of the validation edits accounted for most of the generated error messages attributable to edits in this category.

Apparently, however, many of the more serious validity messages were not cleared up during the edit process. For example, on the school survey, the number of error messages generated by edits indicating missing (not provided) data for teacher and student counts (#16 and 18) remained virtually the same between the time the 1996-97 nonfiscal CCD data were submitted and finalized. Despite this, these particular edits often would have indicated rather unlikely situations (e.g., schools with no teachers or students). The extent to which all error messages were cleaned up during the processing and editing cycle is discussed in more detail in Chapter 5.

Section 4.5 Historical Edits

The historical edit checks produced, by far, the smallest number of error and warning messages generated from the main edit program in the edit phase of the 1996-97 Nonfiscal CCD Surveys. With a combined total of 8 edits generating 2,616 errors, these totals amounted to 9 percent of the edit checks and a mere 3.6 percent of the total error messages.

These edits made comparisons between current year and previous year to check for any data response change. As an example, this would include checking an agency or school type code for a specific record to determine if it had changed between the two surveys.

Only 2 of the edits (#56 and 28 in the agency survey), GRADUATES CY GREATER THAN GRADE 12 PY and GRADS CY WITH NO GRADE 12 STUDENTS PY, with 1,595 and 776 error messages respectively, generated sufficient error messages to account for more than 1 percent of the total. The other six edits together only accounted for slightly more than one-half of one percent of the total error messages generated. The evaluation found this category of edits not to be lacking and, thus, warranting no further investigation or any specific recommendations. Tables 4-4 and 4-5 group the edit checks by type of edit and show the number and percentage of all the error and warning messages generated from the 1996-97 CCD submitted data.

Section 4.6 Use of Edit System Software and 1996-97 CCD Data Error Rates

Review of the *Instructions for Completing the Nonfiscal Surveys of the Common Core of Data 1996-97*[10] manual revealed a comprehensive document covering the states' survey reporting requirements and installation and use of the Edit System. Briefly, the Edit System, available on diskettes or over the internet, is an interactive program that allowed each CCD state coordinator to enter the current year's data, to perform edit checks on entered data, to correct data, to produce edit reports and summary tables of varying detail, and to generate the files that were returned to NCES/agent as initial data submissions. The purpose of providing the CCD Edit System software was to encourage respondents to edit their data prior to submission to NCES or its agent as editing at the source would presumably promote better data quality and processing timeliness.

An important aspect of this effort, then, was to evaluate the effectiveness of the Edit System software. For the purposes of this evaluation, the difference in total errors and error rates for initial data file submissions between the 30 states that used the Edit System software and the 26 states that did not use it[11], was used as a measure of effectiveness. It is cautioned, however, that this measure is limited in its usefulness because other factors, such as the resources available and the ability of each state to implement data collection for the data that make up the CCD surveys, were likely to vary a great deal between the states and could ultimately affect the error counts and the error rates.

This evaluation observed some very interesting findings with regard to the use of the Edit System software. However, the results, at best, were somewhat mixed and inconclusive. Essentially, the underlying assumption that the states which submitted files after using the CCD Edit System would have "cleaner" data files, hence fewer errors and lower error rates, was generally, but certainly not resoundingly, supported. The charts below provide a comparison between the agency universe and school universe surveys based on the total number of errors and the number of errors per record for the states that used the Edit System versus those that did not use it when preparing their initial data submissions.

While the total number of error counts associated with the agency files of states using the Edit System software were less than those for states not using the Edit System, this did not hold true for the school files. This occurrence, as noted below however, was due largely to the high incidence of errors for one state's school data file.

---

[10] *This document is part of the reporting materials which are sent to the states as part of the initial survey mailout.*

[11] *See Table 2-4 in Chapter 2 for a state-by-state listing of the states that did, and did not, use the Edit System.*

Data Submission	No.	Total no. of errors	
		Agency	School
Total*	56	20,591	52,228
States using Edit System software	30	9,425	31,809"
States not using Edit System software	26	11,166	20,419

*Total based on number of records in submitted data files.
"One state, Texas, accounted for 18,641 of this total

While the overall error rate for the school data file was slightly less than one-half the rate of the agency data file, 0.59 to 1.30 errors per record, there were 0.67 fewer errors per record for states using the Edit System than for the states not using the Edit System in the agency files. However, for the school data files the error rate per record was a mere .01 less for the states using the Edit System as opposed to those not using it. The number of errors per record for both the agency and the school surveys are shown below:

Data Submission	No.	No. of errors per record	
		Agency	School
Total*	56	1.30	0.59
States using Edit System software	30	1.04	0.58
States not using Edit System software	26	1.71	0.59

*Total based on number of records in submitted data files.

Similarly, the states' school data files reported, percentage-wise, less than half as many in-error records than did the agency files. The states using the Edit System software for the agency and school data files, respectively, reported 56 and 28 percent of their records in error. These percentages ballooned to about 72 percent for agency data files and 35 percent for school data files with states not using the provided CCD Edit System software when initially submitting data.

Data Submission	No.	% of records in error	
		Agency	School
Total*	56	62.4	30.8
States using Edit System software	30	56.0	28.1
States not using Edit System software	26	71.7	35.0

*Total based on number of records in submitted data files.

The chart below combines the agency and school survey totals showing the overall number of records submitted and the rate of in-error records for the 1996-97 Nonfiscal CCD Surveys:

Data Submission	No.	Agency and School	
		Total no. of records	% of records in error
Total*	56	104,835	35.6
States using Edit System software	30	63,505	32.2
States not using Edit System software	26	41,330	40.7

*Total based on number of records in submitted data files.

Similar values for each individual state are found in Tables 3-2 and 3-3 which were presented at the end of the previous chapter.

Section 4.7 Recommendations

1. Effort should continue to be expended towards convincing more state CCD coordinators to use the Edit System software made available to them for use during data collection and data file preparation. Such effort could be made at the annual data conference training sessions, via the internet, and/or in communications (letter, telephone, fax, etc.) with individual state data coordinators. Based on the findings of this evaluation, at least moderate overall improvement in survey data quality could be realized if more states used the Edit System diskette when

preparing and reporting their initial data submissions.

2. All edits (mostly relational) containing "acceptable" range/tolerance parameters (e.g., current year data value compared to prior year data value exceeded a ±25 percent difference) should be reevaluated for appropriateness and effectiveness. Some edit tolerances, for example, might include an absolute number change criterion or a threshold of comparison could be employed (e.g., if a school has 25 or fewer students then some edits would be suppressed) for some edits. This would ensure a more realistic and useful data criterion "check" aimed at identifying genuine data errors and outliers while at the same time preventing excessive erroneous data failures from being generated. Specific recommendations for individual edits are found in Chapter 5.

3. As suggested in previous survey evaluations, consideration should be given to filling missing or erroneous ZIP code data using commercially available software. This function possibly could be included in the Edit System software revision planned for the upcoming CCD expansion. Although this data element is not crucial, this would eliminate the need for further reviewer and analyst attention.

4. Consideration should be given to adding a "remarks" field to the Edit System software and to state's data base file diskettes which are sent out in the initial mailout. Such a memo field would allow for the recording of pertinent comments and explanations regarding data features or anomalies. This would likely reduce the need for follow-up requiring contact by NCES or its collection agency with the respondent which inevitably lengthens the data processing time. The impact of this suggestion, of course, would be greater if additional respondents were to use the Edit System editing software.

Table 4-1. CCD Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD Agency Universe Survey

This table provides a description of the 60 error and warning messages that can be generated by the CCD main edit program when run on state agency data. Six messages are currently not used.

No.	Type of data problem*	Type of edit*	Error message	No. of error messages generated**	Number of states reporting error
			Total	20,591	-
01	ADM	R	SUPERVISORY UNION ID/ TYPE CODE CONFLICT	0	0
02	CCL	V	BOUNDARY CHANGE CODE INVALID	0	0
03	CCL	H	AGENCY TYPE CODE CHANGED	4	3
04	CCL	H	FIPS CODE CHANGED	22	9
05	CCL	H	METRO STATUS CODE CHANGED	6	4
06	-	-	Not Used	0	0
07	-	-	Not Used	0	0
08	ADM	V	STATE EDUC AGENCY ID IS BLANK OR MISSING	0	0
09	ADM	V	CITY NAME BLANK OR LESS THAN 3 CHARACTER	0	0
10	BAS	R	REGULAR DIPLOMA RECIPIENTS CY/PY > 25%	643	33
11	BAS	V	DROPOUT - NO DATA BY RACE OR GENDER	535	1
12	BAS	V	DROPOUT - NO DATA BY GRADE	1,508	16
13	BAS	V	DROPOUT - NO DATA BY RACE	0	0
14	BAS	V	DROPOUT - NO DATA BY GENDER	0	0
15	BAS	V	DROPOUTS - DATA IN UNKNOWN FIELDS ONLY	162	5
16	BAS	V	AGENCY WITH NO SCHOOLS OR DATA	175	8
17	CCL	V	FIPS COUNTY CODE BLANK OR INVALID	16	2
18	BAS	R	GRADE 12 STUDENTS PY WITH CY GRADS BLANK	0	0
19	BAS	V	INSTRUCTIONAL STAFF DATA NOT PROVIDED	1,219	33
20	CCL	V	AGENCY TYPE CODE INVALID	5	2
21	CCL	V	FIPS CODE NOT VALID FOR STATE	4	2
22	ADM	V	ADDRESS BLANK, M OR LESS THAN 3 CHAR	2	2
23	CCL	V	METRO STATUS CODE INVALID	15	5
24	-	-	Not Used	0	0
25	CCL	R	BOUNDARY CODE NEW BUT HAS NCES ID	0	0
26	ADM	H	AGENCY ID CONFLICT WITH PY	1	1
27	ADM	V	AGENCY NAME INVALID OR LESS THAN 3 CHAR	0	0

Table 4-1. CCD Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD Agency Universe Survey

This table provides a description of the 60 error and warning messages that can be generated by the CCD main edit program when run on state agency data. Six messages are currently not used.

No.	Type of data problem*	Type of edit*	Error message	No. of error messages generated**	Number of states reporting error
28	BAS	H	GRADS CY WITH NO GRADE 12 STUDENTS PY	776	6
29	ADM	V	NCES EDUC AGENCY ID BLANK, M, N, OR 0	0	0
30	BAS	R	OTHER DIPLOMA RECIPIENTS > REG DIPLOMAS	11	5
31	BAS	R	OTHER DIPLOMA RECIPIENTS CY/PY > 25%	24	6
32	BAS	R	OTHER H.S. COMPLETERS > REGULAR DIPLOMAS	10	7
33	BAS	R	OTHER H.S. COMPLETERS CY/PY > 25%	22	12
34	BAS	R	PK-12 STUDENTS CY/PY > 25%	152	26
35	BAS	R	PUPIL/TEACHER RATIO CY/PY > 20%	804	39
36	BAS	R	REG DIPL RECIP CY/ GRADE 12 PY CONFLICT	541	27
37	BAS	R	SCHOOL COUNT CY/PY > 50%	65	29
38	-	-	Not Used	0	0
39	BAS	V	SPECIAL EDUC IEP IS BLANK , 0, M OR N	1,357	40
40	BAS	R	SPEC ED IEP STUDENT COUNT CY/PY > 25%	1,271	26
41	ADM	R	STATE EDUC AGN ID NOT FOUND ON SCH FILE	927	32
42	ADM	R	STATE ABBREVIATION DOES NOT MATCH ZIP	14	3
43	BAS	V	STUDENT COUNTS BLANK, M, N OR 0	543	28
44	BAS	V	SUPPORT STAFF NOT PROVIDED	1,332	39
45	-	-	Not Used	0	0
46	ADM	V	TELEPHONE NUMBER INVALID	0	0
47	BAS	R	TOTAL FTE TEACHER CY/PY > 25%	387	35
48	BAS	R	TOTAL STUDENT COUNT CY/PY > 25%	149	27
49	BAS	V	UNGRADED STUDENTS NOT PROVIDED	72	7
50	BAS	R	UNGRADED STUDENTS CY/PY > 25%	1,146	33
51	ADM	V	ZIP CODE INVALID	465	1
52	BAS	R	STUDENT COUNTS AGENCY/SCHOOL > 25%*	61	20
53	BAS	V	CLASSROOM TEACHERS NOT ON SCHOOL FILE*	779	20

Table 4-1. CCD Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD Agency Universe Survey

This table provides a description of the 60 error and warning messages that can be generated by the CCD main edit program when run on state agency data. Six messages are currently not used.

No.	Type of data problem*	Type of edit*	Error message	No. of error messages generated**	Number of states reporting error
54	BAS	V	SCHOOLS NOT REPORTED ON SCHOOL FILE"	395	21
55	BAS	V	STUDENTS NOT REPORTED BY GRADE - SCH FILE"	174	17
56	BAS	H	GRADUATES CY GREATER THAN GRADE 12 PY	1,595	49
57	BAS	R	SPECIAL ED IEP > 20% OF MEMBERSHIP	611	40
58	BAS	R	SUM OF TEACH/TOTAL FTE > 5%"	0	0
59	BAS	R	TEACHERS ON SCHOOL FILE/TOTAL FTE > 10%	2,591	40
60	-	-	Not Used	0	0

Notes:
*Guide to codes:

Type of edit:
R = Relational
V = Validation
H = Historical

Type of data problem detected by edit check:
ADM = Administrative information
CCL = Code classification
BAS = Basic agency statistics

**Based on data files initially submitted by state CCD coordinators.
"These are warning edits, all others are critical edits.

Table 4-2. CCD Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD School Universe Survey

This table provides a description of the 36 error and warning messages that can be generated by the CCD main edit program when run on state school data. Two messages are currently not used.

No.	Type of data problem*	Type of edit*	Error message	Number of error messages generated**	Number of states reporting error
			Total	52,228	-
01	-	-	Not Used	0	0
02	ADM	H	STATE EDUC AGN ID CHANGED FROM PY	42	7
03	CCL	H	SCHOOL TYPE CODE CHANGED FROM PY	170	17
04	ADM	R	AGENCY NAME DIFFERENT BETWEEN FILES	0	0
05	ADM	V	STATE EDUC AGENCY ID IS BLANK OR MISSING	0	0
06	ADM	V	STATE SCHOOL ID IS BLANK OR MISSING	2	1
07	ADM	V	CITY NAME BLANK OR LESS THAN 3 CHARACTER	3	2
08	-	-	Not Used	0	0
09	CCL	V	OPERATIONAL STATUS CODE INVALID	0	0
10	CCL	V	SCHOOL TYPE CODE INVALID	1	1
11	ADM	V	ADDRESS BLANK, M, LESS THAN 3 CHAR	9	4
12	ADM	V	NCES ID NOT BLANK FOR NEW/ADDED SCHOOL	67	1
13	ADM	V	STATE SCH ID NOT ON PY; OPER STATUS = 1	104	12
14	ADM	V	AGN NAME INVALID OR LESS THAN 3 CHAR	0	0
15	ADM	V	SCHOOL NAME INVALID OR LESS THAN 3 CHAR	0	0
16	BSS	V	STUDENTS NOT REPORTED IN ANY GRADE	2,187	41
17	BSS	V	RACE DATA NOT PROVIDED BUT HAS STUDENTS	668	7
18	BSS	V	CLASSROOM TEACHERS NOT REPORTED	7,593	46
19	BSS	V	TEACHERS AND STUDENTS NOT REPORTED	375	23
20	BSS	R	PUPIL/TEACHER CY/PY RATIO > 20%	12,509	53
21	ADM	R	STATE ABBREVIATION DOES NOT MATCH ZIP	183	9
22	BSS	R	RACIAL CATEGORIES CY/PY > 25%	3,342	53
23	ADM	V	TELEPHONE NUMBER INVALID	332	5
24	BSS	V	SCHOOL WITH 4,000+ STUDENTS	36	6
25	BSS	R	TOTAL STUDENTS BY GRADE CY/PY > 25%	2,253	53
26	BSS	R	TOTAL STUDENTS BY RACE CY/PY > 25%	2,283	53
27	BSS	V	SCHOOL WITH 250+ TEACHERS	4,424	6

Table 4-2. CCD Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD School Universe Survey

This table provides a description of the 36 error and warning messages that can be generated by the CCD main edit program when run on state school data. Two messages are currently not used.

No.	Type of data problem*	Type of edit*	Error message	Number of error messages generated**	Number of states reporting error
28	ADM	V	OPERATIONAL SCHOOL WITH NO VALID NCES ID	0	0
29	ADM	R	SCHOOL NAME = AGENCY NAME	1,040	26
30	ADM	R	STATE AGENCY ID - SCHOOL/AGENCY CONFLICT	3	2
31	ADM	V	ZIP CODE INVALID	2,199	3
32	BSS	V	FREE-LUNCH COUNT IS BLANK"	191	10
33	BSS	R	FREE-LUNCH COUNT > 94.5% MEMBERSHIP"	1,568	38
34	BSS	V	GRADE SEQUENCE GAPS"	2,317	53
35	BSS	R	STUDENT TOTALS BY GRADE/RACE > 25% DIFF"	644	14
36	BSS	R	TEACHER DIFFERENCE CY/PY > 25%"	7,683	49

Notes:
*Guide to codes:

Type of edit:
R = Relational
V = Validation
H = Historical

Type of problem detected by edit check:
ADM = Administrative information
CCL = Code classification
BSS = Basic school statistics

**Based on error data files initially submitted by state CCD coordinators.
"These indicate warning edits, all others are critical edits.

Table 4-3. Main Edit Program Edits and Error and Warning Messages, by Source of Data Problem Detected, 1996-97 Nonfiscal CCD Surveys

This table groups the main edit program error and warning messages according to the type of data problem detected. One category is subdivided into three types of data elements covered. It also shows the number of error messages generated for each category based on the 1996-97 CCD Nonfiscal Surveys submitted data.

Category (Source of data problem detected)	Applicable error/warning message numbers	Total number of edit checks for category*	Number of error messages generated**	
			Number	Percent
Agency Universe				
Total	-	54	20,591	100.0
Administrative information	01, 08-09, 22, 26-27, 29, 41-42, 46, 51	11	1,409	6.8
Code classification information	02-05, 17, 20-21, 23, 25	9	72	0.4
Basic agency statistics (sum of three items below):		34	19,110	92.8
Student counts	10-15, 18, 28, 30-34, 36-37, 39-40, 43, 48-50, 52, 55-57	25	11,428	55.5
Staff counts	19, 44, 47, 53, 58-59	6	6,308	30.6
Agency-wide counts	16, 35, 54	3	1,374	6.7
School Universe				
Total	-	34	52,228	100.0
Administrative information	02, 04-07, 11-15, 21, 23, 28-31	16	3,984	7.6
Code classification information	03, 09-10	3	171	0.3
Basic school statistics (sum of three items below):		15	48,073	92.0
Student counts	16-17, 22, 24-26, 34, 36	8	20,769	39.8
Staff counts	18, 27, 32-33, 35	5	14,420	27.6
School-wide counts	19, 20	2	12,884	24.7

*Two of the error and warning messages for the school universe, and six for the agency universe, are no longer used. Thus, the sum of edit checks for the school and agency universes will equal 34 and 54, respectively.

**Based on data files submitted by state CCD coordinators and generated by the main edit program for the 1996-97 Nonfiscal CCD Surveys.

Table 4-4. Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD Agency Survey, by Type of Edit

This table groups the 54 active main edit program error messages by the type of edit check performed. It also shows the number of error messages generated by each edit from the 1996-97 Nonfiscal CCD submitted data, sorted within category by the number of messages (from highest to lowest).

Edit No.	Edit Description	Number of error messages generated	Percent of error messages generated*
	Total error messages	20,591	100.0
Historical comparisons:	subtotal*	2,404	11.7
56	GRADUATES CY GREATER THAN GRADE 12 PY	1,595	7.7
28	GRADS CY WITH NO GRADE 12 STUDENTS PY	776	3.8
04	FIPS CODE CHANGED	22	0.1
05	METRO STATUS CODE CHANGED	6	(Z)
03	AGENCY TYPE CODE CHANGED	4	(Z)
26	AGENCY ID CONFLICT WITH PY	1	(Z)
Relational comparisons:	subtotal	9,429	45.8
59	TEACHERS ON SCHOOL FILE/TOTAL FTE > 10%	2,591	12.6
40	SPEC ED IEP STUDENT COUNT CY/PY > 25%	1,271	6.2
50	UNGRADED STUDENTS CY/PY > 25%	1,146	5.6
41	STATE EDUC AGN ID NOT FOUND ON SCH FILE	927	4.5
35	PUPIL/TEACHER RATIO CY/PY > 20%	804	3.9
10	REGULAR DIPLOMA RECIPIENTS CY/PY > 25%	643	3.1
57	SPECIAL ED IEP > 20% OF MEMBERSHIP	611	3.0
36	REG DIPL RECIP CY/GRADE 12 PY CONFLICT	541	2.6
47	TOTAL FTE TEACHER CY/PY > 25%	387	1.9
34	PK-12 STUDENTS CY/PY > 25%	152	0.7
48	TOTAL STUDENT COUNT CY/PY > 25%	149	0.7
37	SCHOOL COUNT CY/PY > 50%	65	0.3
52	STUDENT COUNTS AGENCY/SCHOOL > 25%	61	0.3
31	OTHER DIPLOMA RECIPIENTS CY/PY > 25%	24	0.1
33	OTHER H.S. COMPLETERS CY/PY > 25%	22	0.1
42	STATE ABBREVIATION DOES NOT MATCH ZIP	14	0.1
30	OTHER DIPLOMA RECIPIENTS > REG DIPLOMAS	11	0.1
32	OTHER H.S. COMPLETERS > REGULAR DIPLOMAS	10	(Z)
01	SUPERVISORY UNION ID/TYPE CODE CONFLICT	0	0.0
18	GRADE 12 STUDENTS PY WITH CY GRADS BLANK	0	0.0
25	BOUNDARY CODE NEW BUT HAS NCES ID	0	0.0
58	SUM OF TEACH/TOTAL FTE > 5%	0	0.0
Validation checks:	subtotal	8,758	42.5
12	DROPOUT - NO DATA BY GRADE	1,508	7.3
39	SPECIAL EDUC IEP IS BLANK, 0, M, OR N	1,357	6.6
44	SUPPORT STAFF NOT PROVIDED	1,332	6.5
19	INSTRUCTIONAL STAFF DATA NOT PROVIDED	1,219	5.9
53	CLASSROOM TEACHERS NOT ON SCHOOL FILE	779	3.8
43	STUDENT COUNTS BLANK, M, N, OR 0	543	2.6

54

Table 4-4. Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD Agency Survey, by Type of Edit

This table groups the 54 active main edit program error messages by the type of edit check performed. It also shows the number of error messages generated by each edit from the 1996-97 Nonfiscal CCD submitted data, sorted within category by the number of messages (from highest to lowest).

Edit No.	Edit Description	Number of error messages generated	Percent of error messages generated*
11	DROPOUT - NO DATA BY RACE OR GENDER	535	2.6
51	ZIP CODE INVALID	465	2.3
54	SCHOOLS NOT REPORTED ON SCHOOL FILE	395	1.9
16	AGENCY WITH NO SCHOOLS OR DATA	175	0.9
55	STUDENTS NOT REPORTED BY GRADE - SCH FILE	174	0.8
15	DROPOUTS - DATA IN UNKNOWN FIELDS ONLY	162	0.8
49	UNGRADED STUDENTS NOT PROVIDED	72	0.4
17	FIPS COUNTY CODE BLANK OR INVALID	16	0.1
23	METRO STATUS CODE INVALID	15	0.1
20	AGENCY TYPE CODE INVALID	5	(z)
21	FIPS CODE NOT VALID STATE	4	(z)
22	ADDRESS BLANK, M, OR LESS THAN 3 CHAR	2	(z)
27	AGENCY NAME INVALID OR LESS THEN 3 CHAR	0	0.0
02	BOUNDARY CHANGE CODE INVALID	0	0.0
08	STATE EDUC AGENCY ID IS BLANK OR MISSING	0	0.0
09	CITY NAME BLANK OR LESS THAN 3 CHARACTER	0	0.0
13	DROPOUT - NO DATA BY RACE	0	0.0
14	DROPOUT - NO DATA BY GENDER	0	0.0
29	NCES EDUC AGENCY ID BLANK, M, N, OR 0	0	0.0
46	TELEPHONE NUMBER INVALID	0	0.0

Note: *Percentage totals may not add due to rounding.
(z) less than .05 percent.

Table 4-5. Main Edit Program Edits and Number of Error and Warning Messages Generated by Submitted Data in 1996-97 CCD School Survey, by Type of Edit

This table groups the 34 active main edit program error messages by the type of edit check performed. It also shows the number of error messages generated by each edit from the 1996-97 Nonfiscal CCD submitted data, sorted within category by the number of messages (from highest to lowest).

Edit No.	Edit Description	Number of error messages generated	Percent of error messages generated
	Total error messages	52,228	100.0
Historical comparisons:	subtotal*	212	0.4
03	SCHOOL TYPE CODE CHANGED FROM PY	170	0.3
02	STATE EDUC AGN ID CHANGED FROM PY	42	0.1
Relational comparisons:	subtotal	31,508	60.3
20	PUPIL/TEACHER CY/PY RATIO > 20%	12,509	24.0
36	TEACHER DIFFERENCE CY/PY > 25%	7,683	14.7
22	RACIAL CATEGORIES CY/PY > 25%	3,342	6.4
26	TOTAL STUDENTS BY RACE CY/PY > 25%	2,283	4.4
25	TOTAL STUDENTS BY GRADE CY/PY > 25%	2,253	4.3
33	FREE-LUNCH COUNT > 94.5% MEMBERSHIP	1,568	3.0
29	SCHOOL NAME = AGENCY NAME	1,040	2.0
35	STUDENT TOTALS BY RACE/GRADE > 25%	644	1.2
21	STATE ABBREVIATION DOES NOT MATCH ZIP	183	0.4
30	STATE AGENCY ID - SCHOOL/AGENCY CONFLICT	3	(z)
04	AGENCY NAME DIFFERENT BETWEEN FILES	0	0.0
Validation checks:	subtotal	20,508	39.3
18	CLASSROOM TEACHERS NOT REPORTED	7,593	14.5
27	SCHOOL WITH 250+ TEACHERS	4,424	8.5
34	GRADE SEQUENCE GAPS	2,317	4.4
31	ZIP CODE INVALID	2,199	4.2
16	STUDENTS NOT REPORTED IN ANY GRADE	2,187	4.2
17	RACE DATA NOT PROVIDED BUT HAS STUDENTS	668	1.3
19	TEACHERS AND STUDENTS NOT REPORTED	375	0.7
23	TELEPHONE NUMBER INVALID	332	0.6
32	FREE-LUNCH COUNT IS BLANK	191	0.4
13	STATE SCH ID NOT ON PY; OPER STATUS = 1	104	0.2
12	NCES ID NOT BLANK FOR NEW/ADDED SCHOOL	67	0.1
24	SCHOOL WITH 4000+ STUDENTS	36	0.1
11	ADDRESS BLANK, M, LESS THAN 3 CHAR	9	(z)
07	CITY NAME BLANK OR LESS THAN 3 CHARACTER	3	(z)
06	STATE SCHOOL ID IS BLANK OR MISSING	2	(z)
10	SCHOOL TYPE CODE INVALID	1	(z)
05	STATE EDUC AGENCY ID IS BLANK OR MISSING	0	0.0
09	OPERATIONAL STATUS CODE INVALID	0	0.0
14	AGN NAME INVALID OR LESS THAN 3 CHAR	0	0.0
15	SCHOOL NAME INVALID OR LESS THAN 3 CHAR	0	0.0
28	OPERATIONAL SCHOOL WITH NO VALID NCES ID	0	0.0

Note: *Percentage totals may not add due to rounding.
(z) less than .05 percent.

CHAPTER 5. EFFECTIVENESS OF THE NONFISCAL CCD SURVEYS EDIT PROCESS

Section 5.0 Introduction

The error and warning messages generated by the various SAS-based programs during the processing and editing cycle were intended to identify data problems. The stated purpose of this cycle was not primarily to highlight data problems, but rather to prevent them. To serve this purpose effectively, the edit checks needed to be comprehensive enough to identify data problems adversely affecting overall quality, yet be explicit enough so as not to excessively burden the respondents and the analysts who addressed the generated error messages.

This chapter of this evaluation specifically explored the issue of effectiveness of the CCD main edit program edits, and, more generally, the overall effectiveness of the editing and processing procedures of the 1996-97 Nonfiscal CCD Surveys.

Section 5.1 Empirical Results of Initial Data and Final Data Comparison

As described in earlier chapters of this evaluation, the data edit checks for the nonfiscal CCD surveys were applied, to varying degrees, throughout each phase of the processing and editing cycle. A preponderance of these edits was applied during the data summary and main edit stages of the cycle. Although as indicated, some level of review and editing of the data by the state CCD coordinators prior to submission of the agency, school, and state data files to NCES and its agent occurred, the "measurable" editing and correcting that took place occurred during the survey processing cycle described in this report.

To measure the effectiveness of the processing and editing of CCD survey data, the approach used for this evaluation involved applying the CCD main edit program edits to the "initial" data files and the "final" data files. This approach generated initial error and warning messages for the original data as submitted by the state coordinator[12]. The final data files were those files which had been processed through the post-edit stage of the editing and processing cycle and were to be provided to NCES as complete, final data files.

The numbers and types of error and warning messages generated from the final data were compared to those generated from the data initially submitted by the state CCD coordinators. Comparing the error messages, both by edit and by state, resulted in useful information about the quality of the data submissions and the general effectiveness of the edit processing cycle. It is noted that the error and warning messages generated for the initial and final data files were done so by the SAS-based main edit program, modified expressly for this evaluation. That is, application of the main edit program to the data at these two points in the survey cycle was not part of the normal processing procedures. The error message counts presented here as a result do not represent the same error and warning message tallies detailed in Chapter 4.

The intent of this portion of the evaluation was to examine a "before" and "after" snapshot of the states' data files that were submitted, processed and edited. The following summarizes the results of this review:

Totals	Agency universe		
	Final data files	Initial sub-missions	Net difference
Number of records	15,847	15,866	-19
Number of records in error	9,536	10,775	-1,239
Percent of records in error	60.2	67.9	-7.7
Number of errors	18,601	26,030	-7,429

[12] *The original data submissions may have been converted, for compatibility, from a database file into a SAS data set, but no changes to any data were made at this point.*

Totals	School universe		
	Final data files	Initial sub-missions	Net difference
Number of records	88,996	88,965	31
Number of records in error	26,926	32,444	-5,518
Percent of records in error	30.3	36.5	-6.2
Number of errors	49,086	70,315	-21,229

Several noteworthy findings were revealed in this phase of the evaluation. Foremost, the evaluation observed about a one-third overall reduction in the number of errors generated by the edit from the initial to the final data files. For the agency survey, the initial submissions generated 26,030 error warning messages, down to 18,601 errors produced by the final data for a reduction of 7,429 errors (28.5 percent). For the school survey, the same measures were 70,315 errors, down by 21,229 to 49,086 error messages for a decrease of about 30 percent. State-by-state counts are shown in Tables 5-1 and 5-2.

For the agency survey, when evaluated by state, half of the states (28) showed a reduction in the number of generated error messages of less than 10 percent. Nine states reported absolutely no difference in counts between submitted and final data. For only 8 states' submissions was the reduction in errors at least 50 percent. Of these, California and New York together, with a reduction of 4,102 errors, accounted for 55 percent of the total reduction. The remaining 6 states combined accounted for only 167 (2 percent) of the total reduction in error messages.

Closer review of the Table 5-2, the school survey data by state, revealed that only four states[13] accounted for over three-fourths of the overall reduction in error messages. For 27 states, the decline in the numbers of error/warning messages was less then 10 percent. An apparent anomaly present in these figures, however, was that for six states (four in the agency survey and two in the school survey) the number of errors generated actually increased from the initial to the final data checks. A more detailed exploration of data anomalies is presented in a later section of this chapter.

Based on the numbers presented in Tables 5-1 and 5-2, it appeared that after the entire editing process some 70 percent of the generated error messages remained uncorrected. This notion, of course, assumed that all initially generated error messages indicated genuine errors that needed to be corrected. Indeed, one conclusion (or explanation) might be that respondent review determined no need for correction and that the 70 percent of supposed errors were not really "errors" at all. While this extreme may be unlikely, these results suggested a certain level of ineffectiveness in the processing system.

An opposing conclusion might simply be that the editing process was woefully inadequate in that 7 out of every 10 true errors did not get corrected as a result of being subjected to the editing process. Likely the reality was somewhere in the middle ground of these two conclusions. Nonetheless, the issue was a critical one certainly worth additional investigation. A more detailed evaluation of selected edit range and tolerance criteria will be presented in Section 5.4 of this chapter.

As shown in the charts below, a comparison of the error and warning messages generated by the initial submissions to those generated by the final data, by type of edit, revealed that the historical edits showed the biggest percent decline for both the agency and school surveys. This event was likely somewhat skewed because of the relatively small numbers involved, particularly for the school survey. The relational edits showed the least decline for the agency data, while for the school data, the collection of validation edits exhibited the least decline.

[13] *Massachusetts, New York, Tennessee, and Puerto Rico accounted for 16,225 of the total 21,229 error message reductions.*

Type of edit	Agency universe - error/warning messages			
	Final	Initial	Difference	
			No.	%
Total	18,601	26,030	-7,429	-28.5
Relational	8,933	10,492	-1,559	-14.9
Validation	8,081	12,874	-4,793	-37.2
Historical	1,587	2,664	-1,077	-40.4

Type of edit	School universe - error/warning messages			
	Final	Initial	Difference	
			No.	%
Total	49,086	70,315	-21,229	-30.2
Relational	31,215	46,422	-15,207	-32.8
Validation	17,656	23,508	-5,852	-24.9
Historical	215	385	-170	-44.2

Further evaluation in this phase involved analyzing the data to determine if any discernable patterns or links existed, by category or individually, among the edits and associated error and warning messages. Tables 5-3 and 5-4 contain the initial and final error message tallies for the two surveys by edit number. Similar to the comparisons by state presented in Tables 5-1 and 5-2, these two tables revealed several remarkable percent differences for many of the edits.

Over 40 percent (35 of 82) of the edits reported a percent change between 75 to 100 percent. The following chart categorizes the edits by the percent change in the number of error and warning messages generated by the final agency and school data files. Excluded were any edits whose initial submission generated no error messages.

Percent change in number of error and warning messages*	Number of edits	
	Agency	School
Total*	49	33
75 to 100 percent . . .	18	17
50 to 74.9 percent . . .	4	0
25 to 49.9 percent . . .	6	6
0.1 to 24.9 percent . .	20	9
No change	1	1

*Percent change is the absolute change (up or down). For the agency universe five edits, and for the school universe one edit, did not generate error messages for initial submissions and, thus, were omitted .

A closer review of Table 5-3 revealed that the edits demonstrating the largest percentage changes reported the fewest number of error messages to begin with. That is, the edit process made the least difference in those edits generating the greatest number of initial error messages. For the agency survey, the number of error messages generated by edits where change was greater than 50 percent accounted for only 15 percent of all the errors in the initial submissions. Ten edits had all of their associated errors messages corrected during the process, but these only accounted for less than 5 percent of all initial errors. The 21 edits where change was less than 25 percent, on the other hand, accounted for nearly 60 percent of all messages in the initial submissions.

Almost identical percentages to these exist for the school data displayed in Table 5-4. In both surveys, several edits surprisingly generated an increased number of error/warning messages for the final data files in comparison to the initial submissions. As previously mentioned, apparent data anomalies will be addressed later in this chapter.

Section 5.2 Implications for Individual CCD Nonfiscal Records

Based on evaluation of the initial data submissions conducted in this phase of the report, close to 68 percent of agency records and 37 percent of the school records initially submitted for the 1996-97 Nonfiscal CCD Surveys contained errors that potentially would require revision. The submitted data files for the survey cycle evaluated contained a total of 104,831 records - 15,866 in the agency survey and 88,965 in the school survey. While these initial records generated a

combined 96,345 total errors, the number of errors for any individual record ranged from zero to 19. This range for the final data files was between zero and 12 errors. Only 6 records (all on the agency survey) out of the 104,743 final data file records generated as many as 12 error messages.

A distribution of the education agencies and schools by number of error/warning messages revealed that the vast majority (about 95 percent) of records contained 3 or fewer error/warning messages in both the initial data submissions and the final data files.

Number of error and warning messages	Agency universe - number of records in:	
	Final data files	Initial submissions
Total	15,847	15,866
More than 12	0	6
10 to 12	6	74
7 to 9	278	440
4 to 6	648	1,117
1 to 3	8,604	9,138
None	6,311	5,091

Number of error and warning messages	School universe - Number of records in:	
	Final data files	Initial submissions
Total	88,996	88,965
More than 12	0	11
10 to 12	0	8
7 to 9	7	22
4 to 6	1,704	4,213
1 to 3	25,215	28,190
None	62,070	56,521

Section 5.3 Exploring Data Anomalies

As alluded to earlier in this chapter, there were several apparent data anomalies when comparing the states' initial data submissions and final data files. This section highlights some possible explanations to these somewhat disturbing occurrences.

As reported in Tables 5-1 and 5-2, six states (Arkansas, Connecticut, Georgia, and Wyoming on the agency survey, and Washington and Guam on the school survey) reported an increase in generated error and warning messages from their initial to final data files. Although these increases were rather small, totaling 87 and ranging from just 3 to 50 error messages for any given state while averaging less than a 20 percent increase, this anomaly warranted further investigation.

One possible explanation for this event perhaps was the fact that five (excluding Washington) of the six states resubmitted data at some point during the processing and editing cycle, perhaps reintroducing previously corrected errors. It is possible that as "resubmissions" the data were not as rigorously scrutinized as they had been initially. Also, 3 of the states, Arkansas by 17, Washington by 34, and Guam by 33, demonstrated an increase in total records submitted. This may, at least in part, help to explain the increase in the number of errors generated between the initial and final data files. These are but two possible explanations, however, for the apparent anomalies.

Perhaps more disturbing, as depicted in Tables 5-3 and 5-4, 13 data edits (#03, 16, 21, 31, 35, 47, 53, 56, and 57 for the agency survey and #16, 18, 24, and 33 for the school survey) generated more error and warning messages from the final data files than from the initially submitted data files. While the numbers were generally small (8 of the 13 increases represented less than a 10 percent increase), the very event suggested that the edit criteria needed further review.

The 4 edits associated with the school survey where the number of errors increased combined to generate a total of 346 more error messages. This increase was partially explainable by an increase of 31 school records from the initial to the final data files. However, this did not explain the 13 increased edit counts for the agency survey because the increase in the number error messages generated by the final data was associated with 19 fewer agency records.

The edits were further examined in an attempt to determine if any pattern existed for those edits which generated an increased number of error messages. Although no clear pattern was established, the following observations were made regarding the 13 edits which generated more error messages from the final data than from the initial data:

- For the agency data, one edit, CLASSROOM TEACHERS NOT ON SCHOOL FILE (#53), accounted for 281, or over 60 percent, of the 449 increase in errors.

- One school survey edit (#18 CLASSROOM TEACHERS NOT REPORTED) accounted for nearly one-half of the increase reported for that survey.

- By type of edit category, 6 of the 13 were validation edits, 5 were relational edits, and 2 were historical edits, with the validation edits accounting for the majority (nearly 70 percent) of the total increase.

Section 5.4 Tolerance/Range Implications for Selected Main Program Edits

Questions central to issues regarding the processing and editing cycle remain: What data elements generate the greatest number of problems and errors, are the edit criteria appropriate, and what percentage of generated errors are ever corrected? In an attempt to address these issues, this evaluation examined the edits from a more narrow perspective.

Table 5-5 lists the edits which generated the highest total number of error and warning messages from the initially submitted agency and school data and also identifies to what extent they were corrected before the data files were finalized. For both surveys, the top six edits in terms of total number of messages generated, are listed. All of the listed edits accounted for at least five percent of the total number of generated error and warning messages. The edits represented all three types of data edit categories - historical, relational, and validation - described in Chapter 4.

For the agency survey, the top 6 error-generating edits accounted for just more than half (13,673 out of 26,030) of all initial submission-generated error messages. About one in every five errors was generated because instructional or support staff data were not provided by the respondents. The errors associated with this subgroup of edits were, on average, corrected (or reduced) slightly more often - 31.1 to 28.5 percent - than were the agency edit errors as a whole group. On the other hand, the top 6 school universe-related error producing edits accounted for about 70 percent of the total generated error messages. The edits which potentially identified the most data discrepancies and generated the most error messages, particularly for the school survey data, were not corrected, on average, as often as the entire group of school errors were corrected.

These findings suggested that the relatively few edits that generated the majority of errors for both the surveys' data were corrected at a rate below the overall correction rate of 30 percent. Stated somewhat differently, a vast majority of the generated errors, if reviewed at all by the state, were not corrected or, perhaps, did not actually require any modification. Either way, an unnecessary respondent and NCES/agent analyst burden of having to review and filter through large amounts of data apparently may have resulted from ineffective and/or inappropriate data edit criteria parameters. Ideally, data edits should be designed to flag genuine errors while producing a minimum of messages for review. Of particular note, excessive numbers of "false positives" hindered the review process by taxing the efforts of reviewers, thereby increasing the likelihood that legitimate error messages were missed or not fully addressed.

Examination of Table 5-5 reveals that the edits which generated the most errors generally represented two basic types. The types were, one, those associated with situations where data were missing or not provided, and two, situations where data response (or data calculation based on a given response) fell outside of a predetermined value tolerance/range check performed by the CCD main edit program. This second type of error included, for example, cases in which the reported individual racial categories, or total students by race, for the current year and prior year differed by more than ± 25 percent. These general groupings of edits were key because they highlighted two distinctive, but relevant sources of data errors. That is, they distinguished data problems (or data errors) generated by the lack of response to particular survey questions from those errors generated because a data response value did not fall within an acceptable, predetermined range, ostensibly established to identify data outliers. The first group indicated a state-level issue, perhaps a data collection or availability concern, while the second group suggested data problem - likely a data validity or edit check criterion issue.

All edits for the combined agency and school surveys, which generated error messages because data were not provided, potentially either blank, missing, or otherwise invalid, are listed in Table 5-6. Together, these edits accounted for about 22 percent of all errors found in the initial submissions. There was an almost 40 percent reduction in number of errors from the initial to final agency data files. For the school data files, however, this reduction in errors was just under two percent.

Despite the rather uneven reduction in errors between the two surveys, the prevalence of errors generated as a result of missing and invalid data was quite clear. It should be recognized, however, that much of this "non-response" stemmed from the fact that some states, due

to state-level collection requirements, could not provide certain data items, or at least could not provide them according to CCD definitions. Apparently a recurring problem, this ultimately could have been one of the reasons for the low correction rates associated with some of the error counts detailed in chapter 3. Review of the particular edits in Table 5-6 determined them to be thorough and comprehensive, therefore no specific recommendations were made for these edits. A more general approach to address the issue of missing and blank data responses is presented at the end of this chapter.

Table 5-7 shows the 23 edits which checked the submitted data to determine if a reported data figure, or a calculated value based on reported data, fell outside a given tolerance/range parameter. While these edits accounted for about 60 percent of all generated errors for the agency and school data, the reduction in error counts for the final data files was only slightly more than 25 percent. The edits listed in Table 5-7 were of additional interest because they represented the subset of edits that highlighted potential data problems by flagging data responses (and subsequent calculations) which were considered to be outliers or "out-of-range," in other words, data that were potentially accurate but warranted review or verification based on the employed edit criteria. This was in distinct contrast to errors of data omission portrayed in the previous table. Regardless, review and possibly adjustment of certain tolerance/range parameters of the edit criteria would seem advisable.

The issue of proper edit design was relevant in the CCD edit system because the acceptable range/tolerance parameter for some of the edits could likely be revised to reduce the number of messages being generated during the edit process. It is believed that this could be accomplished without sacrificing data quality. This evaluation found that some applied conditions resulted in many responses (data elements) falling outside of acceptable range parameters but which did not result in data corrections (i.e., reduction in error counts).

The error reduction levels previously cited indicated that the edit criteria range/tolerance parameters likely generated excessive edit failures and error messages. For example, on the agency survey, edit #59, which was generated when the total number of teachers reported in schools associated with a given agency exceeded 10 percent (plus or minus) of the total teacher number (in FTE) reported on that agency record for the whole state, initially generated some 3,350 error message flags (13 percent of the total). The final data files indicated a less than one-quarter reduction in number after processing. In the school survey, 2 edits (#20 PUPIL/TEACHER CY/PY RATIO > 20% and #36 TEACHER DIFFERENCE CY/PY > 25%), accounted for over two-fifths of all the error messages, yet the error rate reduction between initial data submissions and final data files for this group of edits was slightly less than 25 percent.

Furthermore, the initial data submissions for the school survey also indicated a total of 4,418 schools with 250 or more teachers (that is, 4,418 generated error messages associated with edit #27 and reported by 6 states). The final data files indicated that no changes or corrections were made to these data responses. Conversely, however, in some instances the selected edit criteria parameters apparently were not sufficiently restrictive. As an example, edit #24 (schools with 4,000 or more students) was generated for only 36 records (schools) - a minuscule 0.04 percent of the reported data[14]. Edit #58 (the sum of individual teachers reported compared with the total FTE reported exceeded a 5 percent difference) in the agency survey was not generated by a single record in the initial data submissions.

The findings presented in this section seemed to support earlier contentions. Simply stated, the apparent error-reduction levels (or lack thereof) between initial and final data files were likely due to the combination of possibilities that the error messages which were generated by edit checks did not truly represent erroneous data (or even data outliers) and therefore actually needed no correction, and/or that the data never were subjected to review by the respondents during the cycle.

In general, the CCD main edit program edits applied to the agency and school data files were found to be very thorough. Section 5-6 provides some general recommendations relating specifically to the edits and edit criteria critiqued in this chapter.

Section 5.5 Impact of Processing and Editing on Selected Final Data Category Counts

As demonstrated, this evaluation observed the processing and editing activities to result in an error correction rate of about only 30 percent. This meant that a full 70 percent of the errors still remained uncorrected or, perhaps never really needed correcting,

---

[14] *Based on the 88,965 school records in the initial data submissions. The increase by 12, to 48 in the final data files, of the number of schools that generated this error was apparently the result of the 31 record increase between the initial and final data files.*

on the final data file. However, knowing what difference the "corrected" 30 percent made in final data counts, for example, might be useful in targeting certain processes and edits that were particularly productive.

To this end, this phase of the evaluation more closely examined several of the total data category counts associated with the initial and final data files. Assessing the differences in final data category total counts versus initial file total counts it was hoped would help in determining the impact, if any, that the processing and editing activities had on various final data counts.

Table 5-8 shows the results for eight selected data categories. Although certainly not comprehensive in nature, this sampling of data category total counts suggested a somewhat moderate impact, with no final total count changing by as much as 15 percent from the initial total count. Thus, while not demonstrating an overly remarkable effect on final data counts, the results of applying the editing process did suggest a certain level of data consistency throughout the cycle. Additional data reliability issues associated with the nonfiscal CCD surveys might be worthy of a future study.

Section 5.6 Recommendations

For the most part, the following recommendations are intended to assist in reducing the large number of generated error and warning messages, particularly those not highlighting true errors, that had to be investigated during the 1996-97 Nonfiscal CCD Surveys. To this end, many of the following suggestions are intended to apply less restrictive checks to the data - that is, to make it more difficult for "good" data to fail a given edit check.

1. Recommendations specific to the agency universe edits include the following:

    a. Edit #35 could be modified to include an absolute change criterion, e.g., by adding an "and" criterion to include a check for an absolute value change of ±5 between the current and prior years' calculated value.

    b. Consideration should be given to modifying 5 edits (#52, 37, 47, 59, and 58) to include an absolute change criterion. This modification should be carried out, at least partly, in order to distinguish between small and large states in terms of the number of students, teachers, education agencies, and schools. Both relative and absolute criteria could be used in combination to qualify the acceptable range for the data amounts reported. The tolerances could vary depending on the size of the reporting unit (e.g., school or agency). This might require the use of an absolute change as the acceptable edit for one size group, a relative change as the acceptable edit tolerance for a different size group, or a combination of both in the edit criteria.

2. The recommendations for the school universe edits are as follows:

    a. Consideration should be given to adding a relative change criterion to two edits (#24 and 27). For example, the edits could be modified with an "and" criterion to include a check for a relative change of ±20 percent change from the previous year. This is similar to the existing edit criteria for several edits (#22, 25, 26, and 36).

    b. Edit #20 could be modified to include an absolute change criterion. One possibility, for example, would be by adding an "and" criterion to include a check for an absolute value change of ±5 between the current and prior years' calculated pupil/teacher ratio. Given the scenario when the pupil/teacher ratio increased from 11 to 14 from one year to the next, for instance, the percentage change would be 27 percent and would elicit an error message. Under the proposed modification, however, both criteria would not be met, thus no error message would result.

    c. Consideration should be given to modifying edit #35 to include an absolute change criterion in order to reduce the number of error and warning messages that do not flag genuine data problems.

3. There should be closer scrutiny of the error and warning messages that indicate missing (not provided) data. These edits, which contributed substantially to the total number of generated error messages for the reviewed surveys, are listed in Table 5-6. In some cases these messages flag true or genuine data problems which should be "correctable" (i.e., missing data would be supplied) during the edit process. There are other recurring situations where states cannot provide a certain type of data item. Historically, NCES and

collection agent are aware of this fact and it is annotated in the "state notes" which accompany the final documentation. One alternative in cases where it is known that data will not be provided (i.e., blank or missing) would be to modify, or code, these situations into the existing main CCD edit program. Alternatively, an edit program based on the state notes could be developed which would suppress error messages for states where the inability to supply certain data has been acknowledged.

4. All edits, for both surveys, that contain range/tolerance criterion (listed in Table 5-7) should be fully scrutinized for appropriateness of the selected tolerance level. However, simply altering a given range or tolerance level, while obviously affecting the number of data errors or "failures" generated, would not speak directly to the central issue of the number of error messages initially generated that ultimately would be corrected during processing and editing procedures. Given the existing survey processing procedures and available documentation, such an analysis was not feasible. It is recommended that this issue be addressed in a future investigation.

Table 5-1. Error and Warning Messages Generated by Initial Submissions Versus Final Data Files, by State, 1996-97 CCD Agency Universe Survey

This table compares the number of main edit program error and warning messages generated by the initial data submissions from the state data coordinators to those from the final 1996-97 CCD data file.

State	Final data files	Initial submissions	Net difference	
			Number	Percent*
All states	18,601	26,030	-7,429	-28.5
Alabama	46	51	-5	-9.8
Alaska	37	146	-109	-74.7
Arizona	588	794	-206	-25.9
Arkansas	174	124	50	40.3
California	563	3,238	-2,675	-82.6
Colorado	65	76	-11	-14.5
Connecticut	82	79	3	3.8
Delaware	24	48	-24	-50.0
District of Columbia	2	2	0	0.0
Florida	53	57	-4	-7.0
Georgia	43	39	4	10.3
Hawaii	0	1	-1	-100.0
Idaho	30	30	0	0.0
Illinois	1,913	1,959	-46	-2.3
Indiana	128	136	-8	-5.9
Iowa	319	319	0	0.0
Kansas	107	107	0	0.0
Kentucky	767	1,195	-428	-35.8
Louisiana	65	69	-4	-5.8
Maine	561	565	-4	-0.7
Maryland	9	9	0	0.0
Massachusetts	1,582	2,069	-487	-23.5
Michigan	1,724	2,027	-303	-14.9
Minnesota	794	941	-147	-15.6
Mississippi	106	112	-6	-5.4
Missouri	172	198	-26	-13.1
Montana	609	641	-32	-5.0
Nebraska	750	849	-99	-11.7
Nevada	17	18	-1	-5.6
New Hampshire	462	598	-136	-22.7
New Jersey**	-	-	-	-
New Mexico	28	48	-20	-41.7
New York	401	1,828	-1,427	-78.1
North Carolina	160	160	0	0.0
North Dakota	119	120	-1	-0.8
Ohio	1,214	1,636	-422	-25.8
Oklahoma	747	773	-26	-3.4
Oregon	102	185	-83	-44.9
Pennsylvania	210	213	-3	-1.4
Rhode Island	9	9	0	0.0
South Carolina	45	60	-15	-25.0
South Dakota	172	175	-3	-1.7
Tennessee	327	470	-143	-30.4
Texas	1,394	1,514	-120	-7.9
Utah	61	79	-18	-22.8
Vermont	471	570	-99	-17.4
Virginia	572	770	-198	-25.7
Washington	566	633	-67	-10.6
West Virginia	28	29	-1	-3.4
Wisconsin	88	105	-17	-16.2
Wyoming	37	33	4	12.1
Dept. of Defense	51	83	-32	-38.6
American Samoa	0	1	-1	-100.0
Guam	2	31	-29	-93.5
Northern Marianas	1	1	0	0.0
Puerto Rico	1	4	-3	-75.0
Virgin Islands	3	3	0	0.0

Notes: *Based on change from initial submissions.
**New Jersey did not submit any data files for 1996-97.

65

BEST COPY AVAILABLE

Table 5-2. Error and Warning Messages Generated by Initial Submissions Versus Final Data Files, by State, 1996-97 CCD School Universe Survey

This table compares the number of main edit program error and warning messages generated by the initial data submissions from the state data coordinators to those from the final 1996-97 CCD data file.

State	Final data files	Initial submissions	Net difference	
			Number	Percent*
All states	49,086	70,315	-21,229	-30.2
Alabama	298	332	-34	-10.2
Alaska	261	277	-16	-5.8
Arizona	545	762	-217	-28.5
Arkansas	189	460	-271	-58.9
California	2,220	2,420	-200	-8.3
Colorado	513	864	-351	-40.6
Connecticut	213	255	-42	-16.5
Delaware	76	87	-11	-12.6
District of Columbia	147	150	-3	-2.0
Florida	1,005	1,206	-201	-16.7
Georgia	699	766	-67	-8.7
Hawaii	101	137	-36	-26.3
Idaho	733	761	-28	-3.7
Illinois	908	1,094	-186	-17.0
Indiana	379	402	-23	-5.7
Iowa	321	327	-6	-1.8
Kansas	289	312	-23	-7.4
Kentucky	382	650	-268	-41.2
Louisiana	268	292	-24	-8.2
Maine	166	170	-4	-2.4
Maryland	160	197	-37	-18.8
Massachusetts	2,055	3,971	-1,916	-48.3
Michigan	1,716	2,285	-569	-24.9
Minnesota	1,815	2,156	-341	-15.8
Mississippi	359	398	-39	-9.8
Missouri	675	740	-65	-8.8
Montana	275	287	-12	-4.2
Nebraska	908	934	-26	-2.8
Nevada	166	175	-9	-5.1
New Hampshire	114	235	-121	-51.5
New Jersey**	-	-	-	-
New Mexico	100	102	-2	-2.0
New York	941	8,545	-7,604	-89.0
North Carolina	268	304	-36	-11.8
North Dakota	193	244	-51	-20.9
Ohio	1,020	1,062	-42	-4.0
Oklahoma	367	405	-38	-9.4
Oregon	239	253	-14	-5.5
Pennsylvania	585	618	-33	-5.3
Rhode Island	100	106	-6	-5.7
South Carolina	271	287	-16	-5.6
South Dakota	340	434	-94	-21.7
Tennessee	3,467	6,681	-3,214	-48.1
Texas	18,046	18,641	-595	-3.2
Utah	126	205	-79	-38.5
Vermont	227	288	-61	-21.2
Virginia	2,180	2,274	-94	-4.1
Washington	878	860	18	2.1
West Virginia	131	217	-86	-39.6
Wisconsin	385	467	-82	-17.6
Wyoming	104	119	-15	-12.6
Dept. of Defense	225	671	-446	-66.5
American Samoa	29	34	-5	-14.7
Guam	27	19	8	42.1
Northern Marianas	34	37	-3	-8.1
Puerto Rico	836	4,327	-3,491	-80.7
Virgin Islands	11	13	-2	-15.4

Notes: *Based on change from initial submissions.
**New Jersey did not submit any data files for 1996-97.

Table 5-3. Error and Warning Messages Generated by Initial Submissions Versus Final Data Files, by Edit, 1996-97 CCD Agency Universe Survey

This table compares the number of main edit program error and warning messages generated by the initial data submissions from the state data coordinators to those generated by the final data files.

Edit Number	Final data files	Initial submissions	Net difference	
			Number	Percent*
Total	18,601	26,030	-7,429	-28.5
01	0	37	-37	-100.0
02	0	39	-39	-100.0
03	4	2	2	100.0
04	10	69	-59	-85.5
05	6	9	-3	-33.3
08	0	2	-2	-100.0
09	0	7	-7	-100.0
10	227	641	-414	-64.6
11	534	535	-1	-0.2
12	1,218	1,506	-288	-19.1
13	0	0	0	(x)
14	0	0	0	(x)
15	0	162	-162	-100.0
16	171	163	8	4.9
17	16	222	-206	-92.8
18	0	0	0	(x)
19	1,232	2,346	-1,114	-47.5
20	5	71	-66	-93.0
21	3	0	3	(x)
22	1	20	-19	-95.0
23	15	103	-88	-85.4
25	0	67	-67	-100.0
26	1	191	-190	-99.5
27	0	6	-6	-100.0
28	51	884	-833	-94.2
29	0	174	-174	-100.0
30	2	11	-9	-81.8
31	24	23	1	4.3
32	3	10	-7	-70.0
33	17	22	-5	-22.7
34	158	175	-17	-9.7
35	804	724	80	11.1
36	532	543	-11	-2.0
37	72	116	-44	-37.9
39	1,347	2,246	-899	-40.0
40	1,220	1,227	-7	-0.6
41	928	1,059	-131	-12.4
42	15	27	-12	-44.4
43	552	613	-61	-10.0
44	1,544	2,716	-1,172	-43.2
46	0	201	-201	-100.0
47	393	378	15	4.0
48	153	166	-13	-7.8
49	69	69	0	0.0
50	1,078	1,144	-66	-5.8
51	0	476	-476	-100.0
52	69	158	-89	-56.3
53	795	514	281	54.7
54	389	489	-100	-20.5
55	190	194	-4	-2.1
56	1,515	1,509	6	0.4
57	667	614	53	8.6
58	0	0	0	(x)
59	2,571	3,350	-779	-23.3

Notes: *Based on change from initial submissions.
(x) - not applicable

Table 5-4. Error and Warning Messages Generated by Initial Submissions Versus Final Data Files, by Edit, 1996-97 CCD School Universe Survey

This table compares the number of main edit program error and warning messages generated by the initial data submissions from the state data coordinators to those generated by the final data files.

Edit Number	Final data files	Initial submissions	Net difference	
			Number	Percent*
Total	49,086	70,315	-21,229	-30.2
02	35	163	-128	-78.5
03	180	222	-42	-18.9
04	0	0	0	(x)
05	0	54	-54	-100.0
06	2	14	-12	-85.7
07	3	51	-48	-94.1
09	0	142	-142	-100.0
10	1	84	-83	-98.8
11	5	103	-98	-95.1
12	0	1,769	-1,769	-100.0
13	107	462	-355	-76.8
14	0	58	-58	-100.0
15	0	24	-24	-100.0
16	2,181	2,098	83	4.0
17	666	687	-21	-3.1
18	7,550	7,393	157	2.1
19	375	407	-32	-7.9
20	12,398	16,505	-4,107	-24.9
21	182	278	-96	-34.5
22	3,171	5,233	-2,062	-39.4
23	0	638	-638	-100.0
24	48	36	12	33.3
25	2,204	3,060	-856	-28.0
26	2,198	3,946	-1,748	-44.3
27	4,418	4,418	0	0.0
28	0	334	-334	-100.0
29	1,030	1,089	-59	-5.4
30	1	986	-985	-99.9
31	2	2,232	-2,230	-99.9
32	0	182	-182	-100.0
33	1,931	1,837	94	5.1
34	2,298	2,322	-24	-1.0
35	183	1,205	-1,022	-84.8
36	7,917	12,283	-4,366	-35.5

Notes: *Based on change from initial submissions.
(x) - not applicable.

Table 5-5. Most Frequently Generated Error and Warning Messages by Initial Data Submissions, 1996-97 Nonfiscal CCD Surveys

This table groups the six most generated main edit program error messages by the initial data submissions for both the agency and school universe surveys by total number and as a percentage. It also shows the net difference, both as a number and a percent, between the error messages generated for the initial data submissions and the final data files.*

Edit No.	Edit Description	Initial Submissions		Net difference	
		Number	Percent	Number	Percent#
Agency Universe					
	All Edits -Total	26,030	100.0	-7,429	-28.5
59	TEACHERS ON SCHOOL FILE/TOTAL FTE > 10%	3,350	12.9	-779	-23.3
44	SUPPORT STAFF NOT PROVIDED	2,716	10.4	-1,172	-43.2
19	INSTRUCTIONAL STAFF DATA NOT PROVIDED	2,346	9.0	-1,114	-47.5
39	SPECIAL EDUC IEP IS BLANK, 0, M, OR N	2,246	8.6	-899	-40.0
56	GRADUATES CY GREATER THAN GRADE 12 PY	1,509	5.8	6	0.4
12	DROPOUT - NO DATA BY GRADE	1,506	5.8	-288	-19.1
	Subtotal/average	13,673	52.5	-4,246	-31.1
School Universe					
	All Edits -Total	70,315	100.0	-21,229	-30.2
20	PUPIL/TEACHER CY/PY RATIO > 20%	16,505	23.5	-4,107	-24.9
36	TEACHER DIFFERENCE CY/PY > 25%	12,283	17.5	-4,366	-35.5
18	CLASSROOM TEACHERS NOT REPORTED	7,393	10.5	157	2.1
22	RACIAL CATEGORIES CY/PY > 25%	5,233	7.4	-2,062	-39.4
27	SCHOOL WITH 250+ TEACHERS	4,418	6.3	0	0.0
26	TOTAL STUDENTS BY RACE CY/PY > 25%	3,946	5.6	-1,748	-44.3
	Subtotal/average	49,778	70.8	-12,126	-24.4

Note: *All edits which accounted for at least five percent of the total number of generated error messages from the initial data submissions are included in this table.
#Based on change from initial submissions.

BEST COPY AVAILABLE

Table 5-6. Selected Error and Warning Messages Related to Blank and Missing Data Generated by Initial Data Submissions, 1996-97 Nonfiscal CCD Surveys

This table highlights, by number and percent, edits and error messages associated with data not provided (blank or missing) as generated by the initial data submissions for both the agency and school universe surveys. It also shows the net difference, both as a number and a percent, between the error messages generated for the initial data submissions and the final data files for these edits.

Edit No.	Edit Description	Initial Submissions		Net difference	
		Number	Percent	Number	Percent*
Agency Universe					
	All Edits -Total	26,030	100.0	-7,429	-28.5
44	SUPPORT STAFF NOT PROVIDED	2,716	10.4	-1,172	-43.2
19	INSTRUCTIONAL STAFF DATA NOT PROVIDED	2,346	9.0	-1,114	-47.5
39	SPECIAL EDUC IEP IS BLANK, 0, M, OR N	2,246	8.6	-899	-40.0
12	DROPOUT - NO DATA BY GRADE	1,506	5.8	-288	-19.1
43	STUDENT COUNTS BLANK, M, N, OR 0	613	2.4	-61	-10.0
11	DROPOUT - NO DATA BY RACE OR GENDER	535	2.1	-1	-0.2
17	FIPS COUNTY CODE BLANK OR INVALID	222	0.9	-206	-92.8
29	NCES EDUC AGENCY ID BLANK, M, N, OR 0	174	0.7	-174	-100.0
49	UNGRADED STUDENTS NOT PROVIDED	69	0.3	0	0.0
22	ADDRESS BLANK, M, OR LESS THAN 3 CHAR	20	0.1	-19	-95.0
09	CITY NAME BLANK OR LESS THAN 3 CHARACTER	7	0.0	-7	-100.0
08	STATE EDUC AGENCY ID IS BLANK OR MISSING	2	0.0	-2	-100.0
13	DROPOUT - NO DATA BY RACE	0	0.0	0	(x)
14	DROPOUT - NO DATA BY GENDER	0	0.0	0	(x)
	Subtotal/average	10,456	40.2	-3,943	-37.7
School Universe					
	All Edits -Total	70,315	100.0	-21,229	-30.2
18	CLASSROOM TEACHERS NOT REPORTED	7,393	10.5	157	2.1
16	STUDENTS NOT REPORTED IN ANY GRADE	2,098	3.0	83	4.0
17	RACE DATA NOT PROVIDED BUT HAS STUDENTS	687	1.0	-21	-3.1
19	TEACHERS AND STUDENTS NOT REPORTED	407	0.6	-32	-7.9
32	FREE-LUNCH COUNT IS BLANK	182	0.3	-182	-100.0
11	ADDRESS BLANK, M, LESS THAN 3 CHAR	103	0.1	-98	-95.1
05	STATE EDUC AGENCY ID IS BLANK OR MISSING	54	0.1	-54	-100.0
07	CITY NAME BLANK OR LESS THAN 3 CHARACTER	51	0.1	-48	-94.1
06	STATE SCHOOL ID IS BLANK OR MISSING	14	0.0	-12	-85.7
	Subtotal/average	10,989	15.6	-207	-1.9

Notes: *Based on change from initial data submissions.
(x) - not applicable

Table 5-7. Selected Error and Warning Messages with Range/Tolerance Criterion Generated by Initial Data Submissions, 1996-97 Nonfiscal CCD Surveys

This table highlights, by number and percent, the edits and error messages associated with data's failure to fall within a predetermined range value as generated by the initial data submissions for both the agency and school universe surveys. It also shows the net difference, both as a number and a percent, between the error messages generated for the initial data submissions and the final data files for these edits.

Edit No.	Edit Description	Initial Submissions		Net difference	
		Number	Percent	Number	Percent*
Agency Universe					
	All Edits -Total	26,030	100.0	-7,429	-28.5
59	TEACHERS ON SCHOOL FILE/TOTAL FTE > 10%	3,350	12.9	-779	-23.3
40	SPEC ED IEP STUDENT COUNT CY/PY > 25%	1,227	4.7	-7	-0.6
50	UNGRADED STUDENTS CY/PY > 25%	1,144	4.4	-66	-5.8
35	PUPIL/TEACHER RATIO CY/PY > 20%	724	2.8	80	11.1
10	REGULAR DIPLOMA RECIPIENTS CY/PY > 25%	641	2.5	-414	-64.6
57	SPECIAL ED IEP > 20% OF MEMBERSHIP	614	2.4	53	8.6
47	TOTAL FTE TEACHER CY/PY > 25%	378	1.5	15	4.0
34	PK-12 STUDENTS CY/PY > 25%	175	0.7	-17	-9.7
48	TOTAL STUDENT COUNT CY/PY > 25%	166	0.6	-13	-7.8
52	SCHOOL COUNT CY/PY > 50%	158	0.6	-89	-56.3
37	STUDENT COUNTS AGENCY/SCHOOL > 25%	116	0.4	-44	-37.9
31	OTHER DIPLOMA RECIPIENTS CY/PY > 25%	23	0.1	1	4.3
33	OTHER H.S. COMPLETERS CY/PY > 25%	22	0.1	-5	-22.7
58	SUM OF TEACH/TOTAL FTE > 5%	0	0.0	0	(x)
	Subtotal/average	8,738	33.6	-1,285	-14.7
School Universe					
	All Edits -Total	70,315	100.0	-21,229	-30.2
20	PUPIL/TEACHER CY/PY RATIO > 20%	16,505	23.5	-4,107	-24.9
36	TEACHER DIFFERENCE CY/PY > 25%	12,283	17.5	-4,366	-35.5
22	RACIAL CATEGORIES CY/PY > 25%	5,233	7.4	-2,062	-39.4
27	SCHOOL WITH 250+ TEACHERS	4,418	6.3	0	0.0
26	TOTAL STUDENTS BY RACE CY/PY > 25%	3,946	5.6	-1,748	-44.3
25	TOTAL STUDENTS BY GRADE CY/PY > 25%	3,060	4.4	-856	-28.0
33	FREE-LUNCH COUNT > 94.5% MEMBERSHIP	1,837	2.6	94	5.1
35	STUDENT TOTALS BY RACE/GRADE > 25%	1,205	1.7	-1,022	-84.8
24	SCHOOLS WITH 4000+ STUDENTS	36	0.1	12	33.3
	Subtotal/average	48,523	69.0	-14,055	-29.0

Note: *Based on change from initial data submissions.
(x) - not applicable

Table 5-8. Comparison Between Initial Submissions and Final Data Files, by Category, for Selected Agency and School Total Data Counts, 1996-97 Nonfiscal Surveys

This table shows several agency and school universe national-level data category totals for both initial data submissions and final data files. It also shows the net difference for each aggregate category total compared between these two files.

Survey (File)	Data category	Total Counts#		Net difference	
		Final data files	Initial submissions	Number	Percent*
Agency	Total FTE of Teachers	2,547,896.0	2,301,683.8	246,212.2	10.7
School	Classroom Teachers	2,390,065.1	2,378,030.1	12,035.0	0.5
School	Total Students	44,946,984.0	44,539,802.0	407,182.0	0.9
Agency	Students with IEP	5,045,138.0	4,533,565.0	511,573.0	11.3
School	Students, Free-lunch Eligible	12,515,444.0	13,852,362.0	-1,336,918.0	-9.7
Agency	Graduates, Regular Diploma	2,240,352.0	2,126,550.0	113,802.0	5.4
Agency	LEA Administrators	44,433.0	40,017.2	4,415.8	11.0
Agency	School Administrators	120,617.0	106,056.3	14,560.7	13.7

Note: #These total values do not include New Jersey. *Based on change from initial data

CHAPTER 6. COMPARISON OF STATE NONFISCAL DATA TO AGENCY AND SCHOOL UNIVERSE DATA

Section 6.0 Introduction

The CCD State Nonfiscal Survey provides three types of information: general identification, staffing, and student. General information includes the name, address, and telephone numbers of each state education agency. Staffing information includes full-time equivalency (FTE) counts for instructional staff, guidance counselors, library staff, administrative staff, and other support staff. Student information includes membership counts by grade and race/ethnicity and high school completers counts by race/ethnicity.

The primary purpose of the state survey is to provide basic information, at the aggregate state level, on public elementary and secondary school students and staff for all states, the District of Columbia, and outlying territories of the United States. State aggregate data cover all education agencies providing free public elementary and secondary education within a state.

The three surveys of the 1996-97 CCD Nonfiscal Surveys provided nested information. The education agencies included on the state survey should be the same as those provided to the NCES or its agent on the Agency Universe Survey described earlier in this evaluation. The schools in the school universe were those reflected in the agency universe, and school and agency surveys were those reflected in the data provided as aggregates in the state survey. The counts from all surveys, though not necessarily equal, should at least theoretically be close, with consistent and explainable differences. This phase of the evaluation examined and compared the state nonfiscal data to the agency and school data to determine their similarity.

Relative to the rather involved editing and processing associated with the agency and school universe surveys, the level and degree of processing and editing involved with the state surveys were somewhat limited. Essentially, after initial submission by the states, the state nonfiscal data files were subject to processing and editing at two points in the survey cycle. The state nonfiscal processing and editing cycle was simultaneous to, and in many instances a part of, the agency and school cycle. Similar to the processing and editing done for the agency and school data, the stages of editing state nonfiscal data correspond to SAS-based software programs and the generated reports that were used to review and edit the CCD data.

The two generated reports which provided the basic guideline for editing the state nonfiscal survey data are detailed in the following sections. They include the "Cross-file Consistency Report" and the "Two-year Consistency Report."

Section 6.1 Cross-file Consistency

The cross-file consistency report was one of three reports generated by the main edit program during the main edit stage of the 1996-97 CCD survey cycle. This report provided a comparison of various corresponding data category totals for student and staffing counts between the reported state nonfiscal data and the reported agency and school data. Variation between these aggregate counts was determined and indicated by both numerical difference and percent difference for each data category (listings of the data categories compared between the state nonfiscal survey data files and the two universe surveys are shown in tables at the end of this chapter).

The cross-file report was part of the edit report package that was sent to the respondents for their review, correction, and response. Similar to the data summary report, the cross-file consistency report allowed respondents to view their submitted data in an aggregate form as well as to review and correct any apparent discrepancies or errors found in their original data submissions. Those data categories having large percent differences were of particular interest and were highlighted for review by the respondents.

Table 3-1 at the end of Chapter 3 identified only 13 data changes/corrections as a result of the examination undertaken by the 20 states that provided a formal, written edit response. These numbers suggested, in addition to the quite high non-response rate for this phase remarked on earlier, that the corrections made to the data as a result of this review were relatively minor. These corrections to the submitted data presented on the cross-file consistency report by the respondents represented a very small percentage (1.5 percent) of the data categories that were reviewed[15].

---

[15] *Based on the 43 data categories/elements, 22 categories compared between the state and agency files and 21 categories compared between the state and school files, that were reviewed by the 20 responding states.*

Section 6.2 Two-year Consistency

The two-year consistency report was created and generated for the first time during the 1996-97 Nonfiscal CCD Surveys cycle. It was not sent to the respondent states for review. This report provided a comparison of current year and prior year data for essentially the same data categories that were found on the cross-file consistency report[16]. Apparently no predetermined, formal or structured guidelines existed to determine which submitted data required follow-up by the respondent. For the survey cycle reviewed, NCES generally determined that any data category total value which exhibited a certain difference from the previous year warranted review and input from the respondent as to its validity. Any necessary edit follow-up typically consisted of data verification and/or correction by state CCD coordinators via fax or telephone communication with NCES/agent who in turn made any necessary changes.

Review of available documentation revealed that 25 of the 56 participating states were asked to verify some aspect of their state nonfiscal data submissions because potential data inconsistencies were identified during the NCES/agent review of the two-year consistency report. This process yielded a total of 48 identifiable data changes/corrections. Five states made no changes to the "questioned" data, while two states did not respond at all to the request to verify their initial state nonfiscal data submission.

There were no shortcomings or extraordinary findings associated with this report and phase of the survey cycle. Rather, this report provided another level, and perspective, of review for state nonfiscal survey data which effectively contributed to the overall assessment of survey data accuracy and reliability.

Section 6.3 Effectiveness of the CCD State Nonfiscal Survey Editing Process

Assessment of the overall effectiveness of the state nonfiscal CCD survey editing process was approached in a manner similar to that used in evaluating the processing and editing for the agency and school surveys described in the previous chapter. To measure the effectiveness of the processing and editing of state nonfiscal data, this phase of the evaluation also applied a modified CCD main edit program to the initial data files and to the final data files in order to generate cross-file consistency reports for both sets of data.

The intent of this portion of the evaluation was essentially to gauge how closely reported state nonfiscal data matched corresponding agency and school data. This evaluation observed there to be remarkable similarity and reliability between the data values reported for the three surveys that make up the nonfiscal CCD surveys as indicated by the relatively small variation that existed between reported data figures.

The cross-file comparison was examined from two perspectives. First, the average variation or difference (as a percentage) of all 43 cross-file data category totals, by state, were calculated[17]. The net difference in average variation as a percent between the initial and final data files was then determined, in part, to provide a measure of effectiveness of the editing process. Using the percent difference for each individual data category compared between the surveys to calculate a state "average" percent variation was a somewhat less than ideal statistical measure for this purpose. Nonetheless, this perspective provided a general barometer of data congruence and revealed several noteworthy findings.

Evidence showed that the data category totals for the state nonfiscal data and the agency universe data, based on average aggregate totals, varied by about seven percent for the initial submissions, down to near three percent for the final data files. This indicated a "coming together" or improvement of data variance of slightly more than four percent, apparently the result of the data editing process undertaken. The reported data values for the state nonfiscal data submissions and the school universe data demonstrated an even greater similarity with an average percent variation of 2.6 percent for initial data and a full 1 percent less for the final data.

In the comparison between state nonfiscal and agency universe data category totals, 10 states demonstrated no variation at all for the 21 common data category totals compared. Twenty-one states reported a less than one percent average variation based on initial data submissions. More than half of the participants (29 states) demonstrated no change at all in the amount of average percent variation between the initial submissions and the final data files. Only 3 states

---

[16] *One data category, "High School Equivalency Recipients," appeared on the two-year report but not on the cross-file report.*

[17] *The state nonfiscal to agency universe comparison included 22 data categories, while the state nonfiscal to school universe comparison included 21 data categories. See Tables 6-3 and 6-4 for a listing of the data categories.*

reported average percent variation improvement of 20 percent or greater.

The state nonfiscal to school universe data category total comparison displayed in the cross-file consistency report also demonstrated remarkably similar data values across the two surveys. For their initial data submissions, 39 states exhibited less than a 2 percent average variation between the state nonfiscal and school data. Five of these states reported perfectly matching data figures. Given the small initial variance between the two surveys' data, it was not surprising that only four states recorded decreases in net average percent variation of at least three percent. However, one state actually saw the discrepancy between reported data category values increase from the initial to the final data files.

That the preponderance of states exhibited no improvement in data category totals variance might have been indicative of the limited review of these data by the respondents or, at the very least, certainly the result of very few corrections made to these data when they were reviewed. It should be noted, however, that any change or correction made to the agency or school data throughout the processing and editing of these data files would affect the aggregate data totals and, subsequently, the amount of variation between these two surveys and the state nonfiscal survey. In any event, the occurrence of minimal change in data congruence between the initial and final data files was not surprising. Table 6-1 highlights, by state, the average percent variance between state nonfiscal and both agency and school universe data totals for initial data submissions and for final data files.

Although the figures reported for the state nonfiscal, agency, and school surveys need not have matched precisely, these data purportedly should have represented the same measures and, thus, would not have differed greatly in value. From another view, Table 6-2 shows the amount of discrepancy that existed between the state nonfiscal and the two universe surveys based on cumulative data totals for the initial data submissions and the final data files. This table reveals that while a certain degree of disparity existed between survey data category totals, more than 70 percent of this disparity was eliminated from the initial to final data files. Specifically, comparison of the state nonfiscal to agency survey data showed that 6 states reported no difference in their initial data submissions and that 31 states exhibited no change in cumulative data category variance totals between the two files. This, again, was not overly surprising given the limited review and edit (i.e., no changes/correction made) of these data. Six states, however, displayed an improvement (or coming together) in data category totals of over 90 percent.

For the state nonfiscal to school universe comparison, only 13 states exhibited no change in cumulative data category variation between initial and final data files while 19 states had their data total discrepancy improve by 90 percent or more. Eight of these 19 states actually saw all of the variation eliminated between the initial and final data files. Two states saw an increase in variation of their data between the two surveys.

A second perspective from which to compare state nonfiscal survey data to agency and school surveys data involved examining initial data submission and final file data by individual data category rather than by aggregate state-level data disparity. Review of Table 6-3 suggested a rather wide range of disparity between data categories compared for the state nonfiscal survey and the agency universe survey. Five of the 21 common data categories compared between the two surveys reported an improvement in data match of greater than 90 percent. However, four data category totals, "prekindergarten teachers," "instructional aides," "instructional coordinators and supervisors," and "student support services" actually reported increases in variation between the initial data submissions and final data files.

Table 6-4 shows that the cumulative variation among the various common data categories compared between the state nonfiscal and school survey data demonstrated a more consistent improvement in data match between files. Here, all categories saw a decrease in disparity, ranging from about a 14 percent to an 86 percent improvement in data value congruence. Fifteen of the 21 data categories compared effectively had totals which came closer together by at least 70 percent between the initial and final measures.

The findings of this evaluation provided strong evidence of accurate and valid initial data submissions accompanied by effective processing procedures for the state nonfiscal survey. While there was only a very small percent difference between the data reported on the three surveys, almost three-quarters of this variance between the initial and final data figures was eliminated as a result of the survey editing process. This suggested reliable, comparable data across all three surveys. However, any such conclusion must be tempered somewhat by the quite low respondent response rates to survey follow up which limited the amount of data that were actually reviewed (verified or corrected) between initial data submission and finalization of the data files.

Section 6.4 Recommendation

1. Incorporate the "State Nonfiscal Two Year Consistency" report and the "Cross-file Consistency" report into one report, highlighting the data which the respondents are requested to review or verify (all data exhibiting a 10 percent difference between prior and current years, for example). This action would not only reduce the redundancies of checking certain data more than once, but also would potentially reduce the number of times a state coordinator is queried about his/her data.

Table 6-1. Comparison of State Nonfiscal to Agency and School Data Category Totals Generated by Initial Submissions Versus Final Data Files, by State, 1996-97 Nonfiscal CCD Surveys

This table highlights the average percent variance, by state, of all corresponding data category totals between state nonfiscal and both the agency and school universe surveys for initial data submissions and final data files.

State	State Nonfiscal to Agency Universe			State Nonfiscal to School Universe		
	Average % variation*		Net difference (Percent)	Average % variation*		Net difference (Percent)
	Final data files	Initial submissions		Final data files	Initial submissions	
All states, average	2.86	7.19	-4.33	1.62	2.62	-1.00
Alabama	0.33	18.72	-18.39	0.26	0.27	-0.01
Alaska	0.16	0.16	0.00	0.01	1.72	-1.71
Arizona	0.16	63.62	-63.46	0.40	3.48	-3.08
Arkansas	0.03	5.03	-5.00	0.75	2.77	-2.02
California	1.89	85.54	-83.65	0.01	0.71	-0.70
Colorado	0.01	0.03	-0.02	0.00	0.01	-0.01
Connecticut	6.83	6.83	0.00	0.59	0.59	0.00
Delaware	0.10	0.47	-0.37	0.00	0.81	-0.81
District of Columbia	0.00	0.00	(x)	0.12	0.12	0.00
Florida	0.00	0.00	(x)	0.09	1.46	-1.37
Georgia	18.05	18.05	0.00	0.00	0.00	(x)
Hawaii	0.02	0.02	0.00	7.65	7.83	-0.18
Idaho	0.04	0.04	0.00	0.11	0.11	0.00
Illinois	0.92	0.92	0.00	14.46	15.07	-0.61
Indiana	0.65	0.65	0.00	0.15	0.71	-0.56
Iowa	0.02	0.02	0.00	0.94	0.94	0.00
Kansas	10.70	10.70	0.00	2.50	2.50	0.00
Kentucky	1.17	41.58	-40.41	1.16	6.31	-5.15
Louisiana	5.48	5.48	0.00	0.08	0.08	0.00
Maine	0.17	0.17	0.00	0.43	0.75	-0.32
Maryland	0.00	0.00	(x)	0.00	0.00	(x)
Massachusetts	0.13	0.13	0.00	0.04	1.33	-1.29
Michigan	9.31	9.31	0.00	1.16	2.60	-1.44
Minnesota	5.29	5.29	0.00	6.47	6.47	0.00
Mississippi	1.61	1.61	0.00	10.29	10.62	-0.33
Missouri	1.12	1.12	0.00	0.55	1.64	-1.09
Montana	0.07	0.07	0.00	0.68	1.03	-0.35
Nebraska	15.76	15.82	-0.06	0.00	0.85	-0.85
Nevada	0.02	0.02	0.00	0.28	3.07	-2.79
New Hampshire	9.71	9.56	0.15	0.07	4.59	-4.52
New Jersey**	-	-	-	-	-	-
New Mexico	5.27	10.40	-5.13	1.95	1.95	0.00
New York	0.50	18.42	-17.92	0.04	0.98	-0.94
North Carolina	0.00	0.00	(x)	0.06	0.06	0.00
North Dakota	0.08	0.30	-0.22	0.00	0.12	-0.12
Ohio	35.28	35.28	0.00	1.56	2.30	-0.74
Oklahoma	19.88	19.89	-0.01	1.33	1.33	0.00
Oregon	1.44	4.14	-2.70	1.57	1.57	0.00
Pennsylvania	0.00	0.00	(x)	0.00	0.00	(x)
Rhode Island	0.05	0.05	0.00	0.00	1.10	-1.10
South Carolina	0.01	0.01	0.00	0.46	1.58	-1.12
South Dakota	0.08	0.08	0.00	0.01	0.84	-0.83
Tennessee	0.01	0.01	0.00	10.96	22.96	-12.00
Texas	1.61	1.66	-0.05	0.00	2.20	-2.20
Utah	0.27	0.27	0.00	2.86	2.86	0.00
Vermont	0.54	1.53	-0.99	15.45	17.26	-1.81
Virginia	0.00	0.00	(x)	0.35	1.64	-1.29
Washington	0.02	0.02	0.00	0.00	1.22	-1.22
West Virginia	0.25	0.25	0.00	0.00	1.09	-1.09
Wisconsin	0.32	0.32	0.00	0.00	1.50	-1.50
Wyoming	0.09	0.09	0.00	0.01	0.03	-0.02
Dept. of Defense**	-	-	-	-	-	-
American Samoa	0.00	0.00	(x)	0.00	0.00	(x)
Guam	0.00	0.00	(x)	0.41	0.41	0.00
Northern Marianas	0.00	0.00	(x)	1.05	1.05	0.00
Puerto Rico	0.00	0.00	(x)	0.13	0.00	0.13
Virgin Islands	2.10	2.10	0.00	1.51	1.51	0.00

Notes: *Absolute difference (up or down) of the change for all categories compared between the two surveys; (x)= not applicable.
**New Jersey did not submit any data files for 1996-97; DoD did not submit a state nonfiscal data file for 1996-97.

Table 6-2. Comparison of State Nonfiscal to Agency and School Category Totals Generated by Initial Submissions Versus Final Data Files, by State, 1996-97 Nonfiscal CCD Surveys

This table highlights the variance, by state, of all corresponding data category totals between state nonfiscal and both the agency and school universe surveys for initial data submissions and final data files.

State	State Nonfiscal to Agency Universe				State Nonfiscal to School Universe			
	Cumulative variation*		Net difference		Cumulative variation*		Net difference	
	Final data files	Initial data files	Number	Percent	Final data files	Initial data files	Number	Percent
Total/average	262,530.7	963,383.4	-700,852.7	-72.7	580,557	1,986,555	-1,405,998	-70.8
Alabama	490.0	1,752.0	-1,262.0	-72.0	5,858	5,879	-21	-0.4
Alaska	77.8	77.8	0.0	0.0	18	6,090	-6,072	-99.7
Arizona	129.5	130,710.9	-130,581.4	-99.9	9,481	53,757	-44,276	-82.4
Arkansas	5.6	185.6	-180.0	-97.0	10,824	37,398	-26,574	-71.1
California	5,437.0	534,253.4	-528,816.4	-99.0	3,286	126,147	-122,861	-97.4
Colorado	3.3	32.1	-28.8	-89.7	0	966	-966	-100.0
Connecticut	1,449.1	1,449.1	0.0	0.0	5,121	5,121	0	0.0
Delaware	3.0	46.6	-43.6	-93.6	0	2,132	-2,132	-100.0
Dist. of Columbia	0.0	0.0	0.0	(x)	285	285	0	0.0
Florida	0.0	0.0	0.0	(x)	8,411	106,852	-98,441	-92.1
Georgia	24,141.3	24,141.3	0.0	0.0	0	0	0	(x)
Hawaii	2.0	2.0	0.0	0.0	44,161	44,481	-320	-0.7
Idaho	6.5	6.5	0.0	0.0	194	194	0	0.0
Illinois	6,266.5	6,266.5	0.0	0.	36,912	44,486	-7,574	-17.0
Indiana	928.9	928.9	0.0	0.0	6,701	26,237	-19,536	-74.5
Iowa	13.2	13.2	0.0	0.0	6,477	6,477	0	0.0
Kansas	1,599.9	1,599.9	0.0	0.0	3,126	3,126	0	0.0
Kentucky	1,105.1	6,062.1	-4,957.0	-81.8	40,922	70,712	-29,790	-42.1
Louisiana	4,899.4	4,899.4	0.0	0.0	1,881	1,881	0	0.0
Maine	69.7	69.7	0.0	0.0	1,443	3,146	-1,703	-54.1
Maryland	3.9	3.9	0.0	0.0	0	0	0	(x)
Massachusetts	377.1	377.1	0.0	0.0	1,725	29,148	-27,423	-94.1
Michigan	21,240.0	21,254.0	-14.0	-0.1	63,578	85,846	-22,268	-25.9
Minnesota	1,307.7	1,307.7	0.0	0.0	165,126	165,126	0	0.0
Mississippi	554.9	554.9	0.0	0.0	8,495	13,809	-5,314	-38.5
Missouri	571.8	571.8	0.0	0.0	7,985	40,150	-32,165	-80.1
Montana	2.9	2.9	0.0	0.0	282	2,220	-1,938	-87.3
Nebraska	400.2	500.1	-99.9	-20.0	0	6,603	-6,603	-100.0
Nevada	4.4	4.4	0.0	0.0	1,574	28,179	-26,605	-94.4
New Hampshire	2,218.7	2,097.3	121.4	5.8	60	4,236	-4,176	-98.6
New Jersey**	-	-	-	-	-	-	-	-
New Mexico	1,708.4	1,857.8	-149.4	-8.0	5,721	5,721	0	0.0
New York	1,902.7	32,700.7	-30,798.0	-94.2	4,348	95,788	-91,440	-95.5
North Carolina	0.0	0.0	0.0	(x)	321	321	0	0.0
North Dakota	5.2	75.9	-70.7	-93.1	0	693	-693	-100.0
Ohio	116,594.0	116,594.0	0.0	0.0	5,099	45,336	-40,237	-88.8
Oklahoma	28,031.9	28,074.3	-42.4	-0.2	4,875	4,875	0	0.0
Oregon	958.5	4,221.5	-3,263.0	-77.3	14,924	14,924	0	0.0
Pennsylvania	24.5	34.5	-10.0	-29.0	0	189	-189	-100.0
Rhode Island	4.5	4.5	0.0	0.0	2	3,332	-3,330	-99.9
South Carolina	2.4	2.4	0.0	0.0	9,923	30,105	-20,182	-67.0
South Dakota	6.1	6.1	0.0	0.0	28	2,805	-2,777	-99.0
Tennessee	4.8	4.8	0.0	0.0	30,558	394,198	-363,640	-92.2
Texas	37,682.7	37,995.9	-313.2	-0.8	0	270,924	-270,924	-100.0
Utah	261.4	261.4	0.0	0.0	5,738	5,738	0	0.0
Vermont	348.4	692.7	-344.3	-49.7	2,364	7,442	-5,078	-68.2
Virginia	2.0	2.0	0.0	0.0	58,754	103,306	-44,552	-43.1
Washington	5.6	5.6	0.0	0.0	0	37,734	-37,734	-100.0
West Virginia	34.7	34.7	0.0	0.0	0	8,049	-8,049	-100.0
Wisconsin	1,385.6	1,385.6	0.0	0.0	0	33,732	-33,732	-100.0
Wyoming	5.9	5.9	0.0	0.0	10	138	-128	-92.8
Dept of Defense**	-	-	-	-	-	-	-	-
American Samoa	20.0	20.0	0.0	0.0	0	0	0	(x)
Guam	0.0	0.0	0.0	(x)	80	80	0	0.0
Northern Marianas	0.0	0.0	0.0	(x)	195	195	0	0.0
Puerto Rico	0.0	0.0	0.0	(x)	3,013	0	3,013	(x)
Virgin Islands	232.0	232.0	0.0	0.0	678	246	432	175.6

Notes: *Absolute difference (up or down) of the change for all categories compared between the two surveys;
**New Jersey did not submit any data files for 1996-97; DoD did not submit a state nonfiscal data file for 1996-97; (x) = not applicable

Table 6-3. Comparison Between Initial Submissions and Final Data Files, by Category, for State Nonfiscal and Agency Data, 1996-97 Nonfiscal CCD Surveys

This table highlights the variation in the twenty-two corresponding data category totals between state nonfiscal and agency universe surveys for both initial data submissions and final data files. It shows the net difference for each category of data compared between these two files.

Data category	Cumulative variation		Net difference	
	Final data files	Initial submissions	Number	Percent*
Total	262,530.7	963,383.4	-700,852.7	-72.7
Prekindergarten Teachers	1,153.1	1,112.8	40.3	3.6
Kindergarten Teachers	553.4	17,064.4	-16,511.0	-96.8
Elementary Teachers	14,387.7	154,385.6	-139,997.9	-90.7
Secondary Teachers	15,056.9	78,993.0	-63,936.1	-80.9
Teachers of Ungraded Classes	1,743.7	26,578.8	-24,835.1	-93.4
Total FTE of Teachers	9,679.9	255,643.7	-245,963.8	-96.2
Instructional Aides	5,602.1	3,806.6	1,795.5	47.2
Instructional Coordinators and Supervisors	1,571.6	1,483.2	88.4	6.0
Elementary Guidance Counselors/Directors	870.6	1,900.3	-1,029.7	-54.2
Secondary Guidance Counselors/Directors	877.7	1,323.7	-446.0	-33.7
Total Guidance Counselors/Directors	1,175.6	2,357.0	-1,181.4	-50.1
Librarians	358.5	1,381.8	-1,023.3	-74.1
Library Support Staff	355.0	1,655.7	-1,300.7	-78.6
LEA Administrators	2,363.6	2,379.1	-15.5	-0.7
Administrative Support Staff	2,656.0	2,756.1	-100.1	-3.6
School Administrators	4,404.9	6,134.2	-1,729.3	-28.2
School Administrative Support Staff	2,398.9	21,665.4	-19,266.5	-88.9
Student Support Services	18,922.9	16,261.6	2,661.3	16.4
All Other Support Services Staff	63,154.6	193,606.4	-130,451.8	-67.4
Regular Diploma High School Completers	112,590.0	124,979.0	-12,389.0	-9.9
Other Diploma Recipients	1,130.0	46,098.0	-44,968.0	-97.5
Other High School Completers	1,524.0	1,817.0	-293.0	-16.1

Note: *Based on change from initial submissions.

BEST COPY AVAILABLE

Table 6-4. Comparison Between Initial Submissions and Final Data Files, by Category, for State Nonfiscal and School Data, 1996-97 Nonfiscal CCD Surveys

This table highlights the variation in the twenty-one corresponding data category totals between state nonfiscal and school universe surveys for both initial data submissions and final data files. It shows the net difference for each category of data compared between these two files.

Data category	Cumulative variation		Net difference	
	Final data files	Initial submissions	Number	Percent*
Total	580,557.0	1,986,555.0	-1,405,998.0	-70.8
Ungraded Students	30,785.0	35,796.0	-5,011.0	-14.0
Prekindergarten Students	10,302.0	39,598.0	-29,296.0	-74.0
Kindergarten Students	12,436.0	42,472.0	-30,036.0	-70.7
Grade 1 Students	8,169.0	38,636.0	-30,467.0	-78.9
Grade 2 Students	8,099.0	34,582.0	-26,483.0	-76.6
Grade 3 Students	8,711.0	35,744.0	-27,033.0	-75.6
Grade 4 Students	8,819.0	39,767.0	-30,948.0	-77.8
Grade 5 Students	8,096.0	43,405.0	-35,309.0	-81.3
Grade 6 Students	9,626.0	53,101.0	-43,475.0	-81.9
Grade 7 Students	11,109.0	47,912.0	-36,803.0	-76.8
Grade 8 Students	10,638.0	45,080.0	-34,442.0	-76.4
Grade 9 Students	11,105.0	51,984.0	-40,879.0	-78.6
Grade 10 Students	9,367.0	30,269.0	-20,902.0	-69.1
Grade 11 Students	9,380.0	32,678.0	-23,298.0	-71.3
Grade 12 Students	7,612.0	16,708.0	-9,096.0	-54.4
Total Students	163,186.0	519,724.0	-356,538.0	-68.6
Total Students - American Indian	6,335.0	11,622.0	-5,287.0	-45.5
Total Students - Asian	28,598.0	44,775.0	-16,177.0	-36.1
Total Students - Hispanic	12,967.0	94,963.0	-81,996.0	-86.3
Total Students - Black	28,993.0	117,648.0	-88,655.0	-75.4
Total Students - White	176,224.0	610,091.0	-433,867.0	-71.1

Note: *Based on change from initial submissions.

Appendix A.

Education Agency Universe Error and Warning Messages/Edit Descriptions

Notes:

*Edit message on the Edit System diskette only.
**Edit message on the CCD main edit program only.

C - Critical edits identify inconsistencies in key fields. These are errors that need to be resolved before the final file may be released.

W - Warning edits identify missing data for non-key fields or significant changes in total fields. As many of these errors will be resolved as possible before the final file is released.

Unless otherwise noted with a 'W', all edits are considered to be critical.

If the boundary change indicator code equals '2', the record will not be edited.

ADDRESS blank, M or less than 3 char
The mailing address is blank, "M" (missing) or less than three characters in length.
The following variable will be displayed:
ADDR1 (first 15 characters)

AGENCY ID conflict with PY
This agency is not reported as new and no agency with this ID exists on the prior year file.

AGENCY NAME invalid or less than 3 char
The education agency name is blank, "M", "N" or less than three characters in length.
The following variable will be displayed:
AGNAME1 (first 15 characters)

*Agency records with same IDs
There are two or more records with the same Agency IDs.

AGENCY TYPE CODE changed
The education agency type has changed since last year.
The following variables will be displayed:
TYPE PYTYPE

AGENCY TYPE CODE invalid
Valid codes are 1-7.
The following variable will be displayed: TYPE

Agency with no schools or data
This agency is not associated with any schools and there are zeros or "N"s in all data fields (students, graduates, dropouts and staffing).

BOUNDARY CHANGE CODE invalid
Valid codes are 1, 2, 3 or 4.
The following variable will be displayed:
BOUND

BOUNDARY CODE new but has NCES ID
The boundary change code is 3 (new), and the NCES Agency ID is not blank, "0" (includes zero filled), "M" or "N".
The following variables will be displayed:
BOUND LEAID

CITY NAME blank or less than 3 character
The name of the city is blank or less than three characters in length.
The following variable will be displayed:
CITY1 (first 15 characters)

CLASSROOM TEACHERS not on school file (W)
This agency has no classroom teachers on the school file.
The following variables will be displayed:
TYPE NUMOFSCH

DROPOUT-Data in UNKNOWN fields only
Dropout data for this agency has been provided for the unknown category only.

DROPOUT-No data by GENDER
Dropout data by gender not reported.

DROPOUT-No data by GRADE
Dropout data for each grade (7-12) are blank or "M".

DROPOUT-No data by RACE
Dropout data for racial/ethnic categories are blank or "M".

DROPOUT-No data by RACE or GENDER
No data in any other dropout categories except ethnicity unknown/gender unknown for grades 7-12 (blank or "M").

*DROPOUT fields invalid
One or more dropout fields contain an entry other than numeric, blank, "M" or "N".

FIPS CODE changed

The FIPS county code has changed since last year.

The following variables will be displayed: CONUM PYCONUM

FIPS CODE not valid for state

The first two digits of the FIPS county code (FIPS state code) are not applicable to the state.

The following variables will be displayed: FIPST ST

FIPS COUNTY CODE blank or invalid

The FIPS county code for this record is blank, missing or not five numeric digits.

The following variable will be displayed: FIPSCNTY

**GRADE 12 STUDENTS PY with CY GRADS BLANK

This agency has prior year grade 12 membership but all current year graduate fields are blank.

**GRADUATES CY GREATER THAN GRADE 12 PY

The total number of graduates on the current year file is greater than the total number of 12th grade students on prior year file.

**GRADS CY with no GRADE 12 STUDENTS PY

No prior year grade 12 students for any schools associated with the agency and the current year graduate fields do not contain "N".

INSTRUCTIONAL STAFF not provided

All instructional staff fields are blank, "M", "N" or 0.

*METRO STATUS 3 with CMSA/PMSA/MSA CODE

The metropolitan status code is 3 for this record, and there is a CMSA/PMSA/MSA numeric code (does not include "0" or zero filled).

The following variables will be displayed: MSC CMSA

METRO STATUS CODE changed

The metropolitan status code has changed since last year.

The following variables will be displayed: MSC PYMSC

METRO STATUS CODE invalid

Valid codes are 1, 2 and 3.

Note: CMSA/PMSA/MSA codes are acceptable as zero filled if metro status code = 3.

The following variable will be displayed: MSC

METRO STATUS/ CMSA/PMSA/MSA incompatible

The metropolitan status code is 1 or 2 for this record, and the CMSA/PMSA/MSA code is "0" (includes zero filled).

The following variables will be displayed: MSC CMSA

NCES EDUC AGENCY ID blank, M, N or 0

Boundary change code is "1" or "2" and the NCES Education Agency ID is blank, "M", "N" or "0" (includes zero filled).

The following variables will be displayed: LEAID BOUND

OTHER DIPLOMA RECIPIENTS > REG DIPLOMAS

This agency has more other diploma recipients than regular diploma recipients.

The following variables will be displayed: OTHDIP REGDIP

OTHER DIPLOMA RECIPIENTS CY/PY > 25%

The number of other diploma recipients on the current year file compared with the prior year file exceeds a 25% difference.

The following variables will be displayed: OTHDIP PYOTHDIP

*OTHER DIPLOMA RECIPIENTS invalid

This field contains an entry other than numeric, blank, "M" or "N".

OTHER H.S. COMPLETERS > REGULAR DIPLOMAS

This agency has more other high school completers than regular diploma recipients.

The following variables will be displayed: OTHCOM REGDIP

OTHER H.S. COMPLETERS CY/PY > 25%

The number of other high school completers on the current year file compared with the prior year file exceeds a 25% difference.

The following variables will be displayed: OTHCOM PYOTHCOM

*OTHER H.S. COMPLETERS invalid
This field contains an entry other than numeric, blank, "M" or "N".

*PK-12 student count invalid
This field contains an entry other than numeric, blank, "M" or "N".

PK-12 students CY/PY > 25%
The PK-12 student total on the current year file compared with the prior year file exceeds a 25% difference.
The following variables will be displayed: PK12 PYPK12

PUPIL/TEACHER ratio CY/PY > 20%
The pupil/teacher ratio (FTE) on the current year file compared to the prior year file exceeds a 20% difference.
The following variables will be displayed: CYRATIO PYRATIO

**REG DIPL RECIP CY/GRADE 12 PY CONFLICT
An education agency with grade 12 last year did not issue any regular diplomas this year.

**REGULAR DIPLOMA RECIPIENTS CY/PY> 25%
The number of regular diploma recipients on the current year file compared with the prior year file exceeds a 25% difference.

*REGULAR DIPLOMA RECIPIENTS invalid
This field contains an entry other than numeric, blank, "M" or "N".

SCHOOLS NOT REPORTED ON SCH FILE (W)
This education agency has no schools on the school file.
The following variable will be displayed: TYPE

**SCHOOL COUNT CY/PY > 50%
The number of schools associated with the agency in the current year file compared with the prior year file exceeds a 50% difference.

SPECIAL ED IEP > 20% of membership
The special education IEP student count exceeds 20% of the membership count.
The following variables will be displayed: SPECED CYSTUDEN

SPECIAL EDUC IEP blank, 0, M or N
Special education IEP is blank, "0", "M" or "N" for this agency.
The following variable will be displayed: TYPE

*SPECIAL EDUC IEP invalid
This field contains an entry other than numeric, blank, "M" or "N".

SPEC ED IEP student count CY/PY > 25%
Special education IEP student count in the current year file compared with the prior year file exceeds a 25% difference.
The following variables will be displayed: SPECED PYSPECED

*STAFF fields invalid
One or more fields contain an entry other than numeric, blank, "M" or "N". All decimals must be implied.

*STATE ABBREVIATION does not match ZIP
The state abbreviation does not contain the expected letter abbreviation for the state when matched to the zip code.
The following variables will be displayed: ST ZIP

STATE ABBREVIATION invalid
The state abbreviation for this record is blank, "M" or "N".
The following variable will be displayed: ST

STATE EDUC AGENCY ID is blank or missing
The state education agency ID is blank or "M".

STATE EDUC AGN ID not found on sch file
The state education agency ID does not agree with the respective ID on the school file.
The following variable will be displayed: STID

STUDENT counts blank, M, N or 0
The student counts (ungraded and PK-12) for this education agency are blank, "M", "N" or "0".
The following variable will be displayed: TYPE

STUDENT counts agency/school > 25% (W)
Reported student counts on the agency file compared to the sum of the students in the associated schools exceeds a 25% difference.
The following variables will be displayed: AGNCNTS SCHCNTS

STUDENTS not reported by GRADE-sch file (W)
This agency has no students by grade on the school file.
The following variables will be displayed: TYPE NUMOFSCH

Sum of TEACHERS/TOTAL FTE > 5% (W)
The sum of individual teachers reported compared with the Total FTE reported exceeds a 5% difference.
The following variables will be displayed: SUMOFTCH RPTTEACH

*SUPERVISORY UNION ID blank
The supervisory union ID number for this record is blank.

SUPERVISORY UNION ID/ TYPE CODE conflict
The supervisory union code is blank, "N", or zero filled. If education agency type code = 2 or 3, there should be a supervisory union ID.
The following variables will be displayed: UNION TYPE

SUPPORT STAFF not provided
All support services staff fields are blank, "M", "N" or 0.

TEACHERS on school file/TOTAL FTE > 10%
The total number of teachers reported in schools associated with this agency exceeds 10% of the Total FTE reported.
The following variables will be displayed: SCTCHCNT TOTTCH

TELEPHONE NUMBER invalid
Telephone numbers must be ten numeric digits, "M" or "N".
The following variable will be displayed: PHONE

TOTAL FTE teacher CY/PY > 25%
Total FTE teacher count on the current year file compared with the prior year file exceeds a 25% difference.
The following variables will be displayed: TOTTCH PYTOTTCH

TOTAL STUDENT COUNT CY/PY > 25%
Total student count on the current year file compared with the prior year file exceeds a 25% difference.
The following variables will be displayed: CYSTUDEN PYSTUDEN

UNGRADED students CY/PY > 25%
Ungraded students on the current year file compared with the prior year file exceeds a 25% difference.
The following variables will be displayed: UG PYUG

*UNGRADED students invalid
This field contains an entry other than numeric, blank, "M" or "N".

ZIP CODE invalid
The zip code is not a five-digit or nine-digit numeric code or is zero filled. It is acceptable for field positions 6-9 to be blank.

Appendix B.

School Universe Error and Warning Messages/Edit Descriptions

Notes:

*Edit message on the Edit System diskette only.
**Edit message on the CCD main edit program only.

C - Critical edits identify inconsistencies in key fields. These are errors that need to be resolved before the final file may be released.

W - Warning edits identify missing data for non-key fields or significant changes in total fields. As many of these errors will be resolved as possible before the final file is released.

Unless otherwise noted with a 'W', all edits are considered to be critical.

If the operational status code equals '2', the record will not be edited.

ADDRESS blank, M or less than 3 char
The mailing address is blank, "M" (missing) or less than three characters in length.
The following variable will be displayed: ADDR1 (first 15 characters)

AGENCY NAME different between files
The education agency name on the school file does not agree with the respective name on the education agency file.
The following variables will be displayed: AGNMESC1 (first 15 characters) AGNMESC2 (last 15 characters) AGNMEAG1 (first 15 characters) AGNMEAG2 (last 15 characters)

AGENCY NAME invalid or less than 3 char
The agency name field is blank, "M", "N" or less than three characters in length.
The following variable will be displayed: AGNAME1 (first 15 characters)

CITY NAME blank or less than 3 character
The name of the city is blank or less than three characters in length.
The following variable will be displayed: CITY1 (first 15 characters)

*CLASSROOM TEACHERS invalid
This field contains an entry other than numeric, blank, "M" or "N". All decimals must be implied.
The following variable will be displayed: FTE

CLASSROOM TEACHERS not reported
This field is blank, "0", "M" or "N".

FREE-LUNCH ELIGIBLE > 94.5% membership(W)
The free-lunch eligible count is greater than 94.5% of all members.
The following variables will be displayed: FLE STUDENTS

*FREE-LUNCH ELIGIBLE invalid (W)
This field contains an entry other than numeric, "M" or "N".
The following variable will be displayed: FLE

FREE-LUNCH ELIGIBLE is blank (W)
The free-lunch eligible count is blank.

*GRADE fields - one or more invalid
One or more of the grade fields contain an entry other than numeric, blank, "M" or "N".

GRADE sequence gaps (W)
Between the lowest and highest grades, one or more grades have no students.

NCES ID not blank for new/added school
The operational status code is 3 (new) or 4 (added), and the NCES school ID is not blank for this record.
The following variable will be displayed: SCHNO

Operational school with no valid NCES ID
Operational status code = 1 and NCES Education Agency ID or NCES School ID is blank, "M" or "N".
The following variable will be displayed: LEAID SCHNO

OPERATIONAL STATUS CODE invalid
Valid codes are 1-4.
The following variable will be displayed: STATUS

PUPIL/TEACHER ratio CY/PY > 20%
The current year pupil/teacher ratio (FTE), in a regular-type school compared to the prior year exceeds a 20% difference.
The following variable will be displayed: CYRATIO PYRATIO

RACE data not provided but has STUDENTS

Students enrolled, but no students reported in any racial category. All fields are blank, "0", "M" or "N". This edit is critical for regular schools and a warning for all other types.

*RACE fields - one or more invalid

One or more of the racial categories contain an entry other than numeric, blank, "M" or "N".

RACIAL categories CY/PY > 25%

Current year individual racial categories are compared to prior year. Based on this comparison, an individual category exceeds a 25% difference. This edit is critical for regular schools and a warning for all other schools.
The following variables will be displayed: CATEG CYRACE PYRACE

SCHOOL NAME = AGENCY NAME

The school name is identical to the agency name.
The following variables will be displayed: AGNAME1 (first 15 characters) AGNAME2 (last 15 characters)

SCHOOL NAME invalid or less than 3 char

The school name is blank, "M", "N" or less than three characters in length.
The following variable will be displayed: SCNAME1 (first 15 characters)

*SCHOOL RECORDS WITH THE SAME ID

There are two or more records with the same Agency and School IDs.

SCHOOL TYPE CODE changed from PY

The school type code has changed since last year.
The following variables will be displayed: TYPE PYTYPE

SCHOOL TYPE CODE invalid

Valid codes are 1-4.
The following variable will be displayed: TYPE

School with 250+ TEACHERS

All schools with 250 or more teachers are listed for review.
The following variable will be displayed: FTE

School with 4,000+ STUDENTS

All schools with 4,000 or more students are listed for review.
The following variable will be displayed: STUDENTS

STATE ABBREVIATION does not match ZIP

The state abbreviation does not contain the expected letter abbreviation for the state when matched to the zip code.
The following variables will be displayed: ST ZIP

*STATE ABBREVIATION invalid

The state abbreviation for this record is blank, "M" or "N".
The following variable will be displayed: ST

*STATE AGENCY ID not on agency file

The school record did not match with a record on the agency file.
The following variable will be displayed: STID

**STATE AGENCY ID-SCHOOL/AGENCY conflict

The state education agency ID for this school does not agree with the respective ID on the education agency file.

**STATE EDUC AGN ID CHANGED FROM PY

The state education agency ID has changed since last year.

STATE EDUC AGENCY ID is blank or missing.

The state education agency ID is blank or "M".

STATE SCH ID not on PY; OPER STATUS = 1

The status code indicates that this school was operational last year, but there is no record on the prior year file with this State School ID. The State School ID and the operational Status Code must be compatible.
The following variable will be displayed: STID

STATE SCHOOL ID is blank or missing

The state school ID is blank or "M".

STUDENT TOTAL BY GRADE CY/PY > 25%

The current year student by grade total compared to the prior year student by grade total exceeds a 25% difference. This edit is critical for regular schools and a warning for all other types.
The following variables will be displayed: CYMEMBER PYMEMBER

STUDENT TOTAL BY RACE CY/PY > 25%

The current year student by race total compared to the prior year student by race total exceeds a 25% difference. This edit is critical for regular schools and a warning for all other types.
The following variables will be displayed: CYTOTAL PYTOTAL

STUDENT TOTALS BY GRADE/RACE > 25%(W)

The total number of students by grade and total number of students by race exceed a 25% difference.
The following variables will be displayed: STUDENTS RACECNTS

STUDENTS not reported in any GRADE

No students reported in any grade (Ungraded - Grade 12). All fields are blank, "0", "M" or "N".

TEACHER difference CY/PY > 25% (W)

The difference of the classroom teacher count from last year to this year exceeds a 25% difference.
The following variables will be displayed: FTE PYFTE

TEACHERS and STUDENTS not reported

No teachers or students reported for this school. All regular schools (type 1) are expected to have student and teacher data.

TELEPHONE NUMBER invalid

Telephone numbers must be ten numeric digits, "M" or "N".
The following variable will be displayed: PHONE

ZIP CODE invalid

The zip code must provide a five-digit or nine-digit numeric code. It is acceptable for field positions 6-9 to be blank.
The following variables will be displayed: ZIP ZIP4

Listing of NCES Working Papers to Date

Please contact Angela Miles at (202) 219-1761 (angela_miles@ed.gov)
if you are interested in any of the following papers

Number	Title	Contact
94-01 (July)	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94-02 (July)	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94-03 (July)	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94-04 (July)	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94-05 (July)	Cost-of-Education Differentials Across the States	William Fowler
94-06 (July)	Six Papers on Teachers from the 1990-91 Schools and Staffing Survey and Other Related Surveys	Dan Kasprzyk
94-07 (Nov.)	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
95-01 (Jan.)	Schools and Staffing Survey: 1994 Papers Presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95-02 (Jan.)	QED Estimates of the 1990-91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95-03 (Jan.)	Schools and Staffing Survey: 1990-91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk
95-04 (Jan.)	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95-05 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors	Jeffrey Owings

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
95-06 (Jan.)	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95-07 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95-08 (Feb.)	CCD Adjustment to the 1990-91 SASS: A Comparison of Estimates	Dan Kasprzyk
95-09 (Feb.)	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95-10 (Feb.)	The Results of the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95-11 (Mar.)	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
95-12 (Mar.)	Rural Education Data User's Guide	Samuel Peng
95-13 (Mar.)	Assessing Students with Disabilities and Limited English Proficiency	James Houser
95-14 (Mar.)	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
95-15 (Apr.)	Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey	Sharon Bobbitt
95-16 (Apr.)	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95-17 (May)	Estimates of Expenditures for Private K-12 Schools	Stephen Broughman
95-18 (Nov.)	An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey	Dan Kasprzyk
96-01 (Jan.)	Methodological Issues in the Study of Teachers' Careers: Critical Features of a Truly Longitudinal Study	Dan Kasprzyk

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-02 (Feb.)	Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association	Dan Kasprzyk
96-03 (Feb.)	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
96-04 (Feb.)	Census Mapping Project/School District Data Book	Tai Phan
96-05 (Feb.)	Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey	Dan Kasprzyk
96-06 (Mar.)	The Schools and Staffing Survey (SASS) for 1998-99: Design Recommendations to Inform Broad Education Policy	Dan Kasprzyk
96-07 (Mar.)	Should SASS Measure Instructional Processes and Teacher Effectiveness?	Dan Kasprzyk
96-08 (Apr.)	How Accurate are Teacher Judgments of Students' Academic Performance?	Jerry West
96-09 (Apr.)	Making Data Relevant for Policy Discussions: Redesigning the School Administrator Questionnaire for the 1998-99 SASS	Dan Kasprzyk
96-10 (Apr.)	1998-99 Schools and Staffing Survey: Issues Related to Survey Depth	Dan Kasprzyk
96-11 (June)	Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with comments on School Reform, Governance, and Finance	Dan Kasprzyk
96-12 (June)	Predictors of Retention, Transfer, and Attrition of Special and General Education Teachers: Data from the 1989 Teacher Followup Survey	Dan Kasprzyk
96-13 (June)	Estimation of Response Bias in the NHES:95 Adult Education Survey	Steven Kaufman
96-14 (June)	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-15 (June)	Nested Structures: District-Level Data in the Schools and Staffing Survey	Dan Kasprzyk
96-16 (June)	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
96-17 (July)	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
96-18 (Aug.)	Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children	Jerry West
96-19 (Oct.)	Assessment and Analysis of School-Level Expenditures	William Fowler
96-20 (Oct.)	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96-21 (Oct.)	1993 National Household Education Survey (NHES:93) Questionnaires: Screener, School Readiness, and School Safety and Discipline	Kathryn Chandler
96-22 (Oct.)	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
96-23 (Oct.)	Linking Student Data to SASS: Why, When, How	Dan Kasprzyk
96-24 (Oct.)	National Assessments of Teacher Quality	Dan Kasprzyk
96-25 (Oct.)	Measures of Inservice Professional Development: Suggested Items for the 1998-1999 Schools and Staffing Survey	Dan Kasprzyk
96-26 (Nov.)	Improving the Coverage of Private Elementary-Secondary Schools	Steven Kaufman
96-27 (Nov.)	Intersurvey Consistency in NCES Private School Surveys for 1993-94	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-28 (Nov.)	Student Learning, Teaching Quality, and Professional Development: Theoretical Linkages, Current Measurement, and Recommendations for Future Data Collection	Mary Rollefson
96-29 (Nov.)	Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2-Year-Olds in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
96-30 (Dec.)	Comparison of Estimates from the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-01 (Feb.)	Selected Papers on Education Surveys: Papers Presented at the 1996 Meeting of the American Statistical Association	Dan Kasprzyk
97-02 (Feb.)	Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-03 (Feb.)	1991 and 1995 National Household Education Survey Questionnaires: NHES:91 Screener, NHES:91 Adult Education, NHES:95 Basic Screener, and NHES:95 Adult Education	Kathryn Chandler
97-04 (Feb.)	Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-05 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-06 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-07 (Mar.)	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97-08 (Mar.)	Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey	Kathryn Chandler

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-09 (Apr.)	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
97-10 (Apr.)	Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993-94 School Year	Dan Kasprzyk
97-11 (Apr.)	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97-12 (Apr.)	Measuring School Reform: Recommendations for Future SASS Data Collection	Mary Rollefson
97-13 (Apr.)	Improving Data Quality in NCES: Database-to-Report Process	Susan Ahmed
97-14 (Apr.)	Optimal Choice of Periodicities for the Schools and Staffing Survey: Modeling and Analysis	Steven Kaufman
97-15 (May)	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97-16 (May)	International Education Expenditure Comparability Study: Final Report, Volume I	Shelley Burns
97-17 (May)	International Education Expenditure Comparability Study: Final Report, Volume II, Quantitative Analysis of Expenditure Comparability	Shelley Burns
97-18 (June)	Improving the Mail Return Rates of SASS Surveys: A Review of the Literature	Steven Kaufman
97-19 (June)	National Household Education Survey of 1995: Adult Education Course Coding Manual	Peter Stowe
97-20 (June)	National Household Education Survey of 1995: Adult Education Course Code Merge Files User's Guide	Peter Stowe
97-21 (June)	Statistics for Policymakers or Everything You Wanted to Know About Statistics But Thought You Could Never Understand	Susan Ahmed
97-22 (July)	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-23 (July)	Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form	Dan Kasprzyk
97-24 (Aug.)	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97-25 (Aug.)	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
97-26 (Oct.)	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
97-27 (Oct.)	Pilot Test of IPEDS Finance Survey	Peter Stowe
97-28 (Oct.)	Comparison of Estimates in the 1996 National Household Education Survey	Kathryn Chandler
97-29 (Oct.)	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Steven Gorman
97-30 (Oct.)	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Steven Gorman
97-31 (Oct.)	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Steven Gorman
97-32 (Oct.)	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questionnaires)	Steven Gorman
97-33 (Oct.)	Adult Literacy: An International Perspective	Marilyn Binkley
97-34 (Oct.)	Comparison of Estimates from the 1993 National Household Education Survey	Kathryn Chandler
97-35 (Oct.)	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
97-36 (Oct.)	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-37 (Nov.)	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Steven Gorman
97-38 (Nov.)	Reinterview Results for the Parent and Youth Components of the 1996 National Household Education Survey	Kathryn Chandler
97-39 (Nov.)	Undercoverage Bias in Estimates of Characteristics of Households and Adults in the 1996 National Household Education Survey	Kathryn Chandler
97-40 (Nov.)	Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey	Kathryn Chandler
97-41 (Dec.)	Selected Papers on the Schools and Staffing Survey: Papers Presented at the 1997 Meeting of the American Statistical Association	Steve Kaufman
97-42 (Jan. 1998)	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
97-43 (Dec.)	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
97-44 (Dec.)	Development of a SASS 1993-94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98-01 (Jan.)	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
98-02 (Jan.)	Response Variance in the 1993-94 Schools and Staffing Survey: A Reinterview Report	Steven Kaufman
98-03 (Feb.)	Adult Education in the 1990s: A Report on the 1991 National Household Education Survey	Peter Stowe
98-04 (Feb.)	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
98-05 (Mar.)	SASS Documentation: 1993-94 SASS Student Sampling Problems; Solutions for Determining the Numerators for the SASS Private School (3B) Second-Stage Factors	Steven Kaufman
98-06 (May)	National Education Longitudinal Study of 1988 (NELS:88) Base Year through Second Follow-Up: Final Methodology Report	Ralph Lee
98-07 (May)	Decennial Census School District Project Planning Report	Tai Phan
98-08 (July)	The Redesign of the Schools and Staffing Survey for 1999-2000: A Position Paper	Dan Kasprzyk
98-09 (Aug.)	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
98-10 (Aug.)	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
98-11 (Aug.)	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96-98) Field Test Report	Aurora D'Amico
98-12 (Oct.)	A Bootstrap Variance Estimator for Systematic PPS Sampling	Steven Kaufman
98-13 (Oct.)	Response Variance in the 1994-95 Teacher Follow-up Survey	Steven Kaufman
98-14 (Oct.)	Variance Estimation of Imputed Survey Data	Steven Kaufman
98-15 (Oct.)	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
98-16 (Dec.)	A Feasibility Study of Longitudinal Design for Schools and Staffing Survey	Stephen Broughman
98-17 (Dec.)	Developing the National Assessment of Adult Literacy: Recommendations from Stakeholders	Sheida White

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
1999-01 (Jan.)	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
1999-02 (Feb.)	Tracking Secondary Use of the Schools and Staffing Survey Data: Preliminary Results	Dan Kasprzyk
1999-03 (Feb.)	Evaluation of the 1996-97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young

DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA